nature

www.nature.com/nature Vol 454 | Issue no. 7200 | 3 July 2008

Feasting and fasting

Bad nutrition needs the world's attention. Not least that of biologists.

'Exterminate the metabolic syndrome!' That was the mission of Japan's government when two vice-ministers for health announced in 2006 that they would cut their waistlines. Encouraged no doubt by the media's keen scrutiny of their progress, they duly did so, with Keizo Takemi's girth coming down from 100.5 to 95 cm over six months.

As was clear from a wide-ranging survey of nutrition issues at last month's Pacific Health Summit in Seattle, Washington, behavioural change is a key element in tackling the twin crises of under-nutrition and obesity, which are all too easily forgotten in the competition for the world's attention.

What is also insufficiently recognized is the dual nature of the challenge. The combination of "feasting and fasting", as Margaret Chan, director-general of the World Health Organization (WHO), described it in Seattle, increasingly arises as countries develop. For example, China and Mexico are showing both aspects of the 'nutrition transition' — from extreme under-nutrition to obesity caused by insufficient exercise and an excess of unhealthy diets, and diabetes and cancers that follow.

Researchers have tracked the effects of under-nutrition on mothers from before they become pregnant and on their children in the first two years of life. Not only does it cause immediate health problems for the offspring, but it affects their health, educational and work prospects for the rest of their lives. (For an influential overview, see the series published by *The Lancet* online in January 2008, http://www.thelancet.com/online/focus/undernutrition.) Those longer-term consequences, including an increase in type 2 diabetes, exacerbate the health problems that arise with obesity as countries progress in their development.

Priorities for action to encourage behavioural change include education, compulsory labelling of the calorie content of restaurant menus and regulation of the food industry. Change is also required in farming and retailing practices for healthier food, in both developed and developing countries.

But science matters too. And, above all, science that spans the full range — the study of diets and exercise habits of populations, of physiology, and of molecular mechanisms and genetics — and that includes the social sciences of behaviour as well as the natural sciences of nutrients. For example, cohort studies that examine the dietary intake of women — so often it is women who play a crucial role in public health and who need empowerment — can be related to physiological studies of the effects of dietary components, which in turn can be related to effects of a particular micronutrient on human or model-organism gene expression. Such a holistic approach has the potential for an immediate impact on policies.

"A holistic approach to under-nutrition has the potential for an immediate impact on policies."

Some funding agencies say that they don't receive strong proposals in nutrition. But researchers reasonably assert that funders have only themselves to blame, having weakened organismal biology in favour of cell and molecular studies for many years. A commendable exception to this is France, which for several decades has built and sustained the science of nutrition in a strategic and multifaceted way.

What can best motivate more researchers to tackle these challenges? For a start, an awareness of the sheer impacts of bad nutrition: 35% of under-five child deaths and 11% of the global disease burden are caused by under-nutrition, with poor breastfeeding and deficiencies in zinc and vitamin A being major factors. Meanwhile a poor energy balance in individuals lies at the heart of the epidemic in obesity, which has been estimated by the WHO to give rise to as much as 7% of global health-care costs in developed countries, including a rising tide of diabetes and cancers.

But ultimately, governments need to take the lead. In April, Japan announced a follow-up to its ministers' personal regime of eating less and moving more with a national programme of health check-ups. Next month it hosts the G8 summit, where the cause of improving nutrition will have its opportunity as the participants focus on the food crisis. Science is a critical element requiring the G8's attention in tackling the huge and global burdens of bad nutrition. ■

In rude health

A treasure-trove of data in the UK National Health Service is set to energize biomedical research.

Reaching 60 is often associated with people slowing down and beginning to feel their age. Britons, for instance, are entitled to a free bus pass. But their universal health-care system, the National Health Service (NHS), which celebrates its 60th birthday on 5 July, is broadly considered to be in better shape now than it was at its half-centenary — even though baby boomers are greyer and new medicines are dearer. Scrapping it and switching to a private system would be unthinkable.

NHS-bashing has become a UK national pastime, and shortcomings are inevitable in what is one of the world's five biggest employers. But Britons who use the service regularly judge it more positively than those who mainly talk about it. British newspapers are often rife with reports of bad management, but a recent survey showed 91% of 17 million hospital inpatients rated their care as good, very good or excellent. And although the NHS sometimes fares badly in surveys that focus on how well specific illnesses are treated, probably the most thorough evaluation of late — a six-country study of equity, efficiency, quality, access, and long and productive patient lives by independent foundation the

Commonwealth Fund — ranked it top, above the health-care systems of Australia, Canada, Germany, New Zealand and the United States.

Scientific research has been a core goal of the NHS since its foundation — but one might be forgiven for not knowing it. For years, the funds for research have been distributed within the regional health-care deliverers in a system that might have been designed to obstruct collaboration with universities and drugs companies. All credit, then, to the NHS's director-general of research and development, Sally Davies. With the creation of the National Institute for Health Research (NIHR), a virtual body within the NHS, Davies has dragged the funds into the daylight. By 2011, these are expected to amount to about £1 billion (US$2 billion).

In its attention to researchers' careers, networks and collaborations, and to transparent indicators of achievement, Davies' Best Research for Best Health programme is helping to transform the research landscape. Plans include virtual organizations to link universities, hospitals and industry; ten 'academic health centres' resembling US university hospitals, through which researchers will be able to conduct studies on patients more easily; and somewhere between 15 and 50 'health innovation and educational clusters', which the government hopes will speed up procedural innovation, promising better funding for academics who get involved. Meanwhile, NIHR collaboration with the Medical Research Council, after a fractious start, is beginning to make progress.

Perhaps the most significant move for biomedical science at the NHS lies in opening up the ocean of patients' data that the organization has collected over the decades. Public consultations are now beginning that, favourable soundings suggest, will lead to ways by which researchers can readily find appropriate patients for research and clinical trials, and can gain access to data whose sources are anonymized but traceable subject to the patient's permission. In particular, the national extent and depth of those data will provide researchers in academia and industry with a globally unique resource for highly targeted studies and clinical trials — a key element of translational medicine.

Such an information system will rightfully evoke worries about privacy. In the future, those concerns will become more sensitive as genetic testing becomes more predictively powerful. Yet, at the same time, as that era blossoms, it will bring the risk-pooling benefits of universal health-care to the fore. As long as people do not have to share genetic data with private insurers, as is the case in the United Kingdom until at least 2014, those who anticipate bad health will do well to buy insurance cover. The genetically lucky, meanwhile, might as well save money and rely on the state. This will squeeze private insurers, suggesting that the NHS's golden period may be yet to come. ■

Does the past have a future in Berlin?

Not unless a research institution's managers recognize its value.

Most museums have so little space that they hold the majority of their collections in storage. But the Berlin Medical Historical Museum has problems that border on the absurd. Its overflow is piled up in entirely unsuitable make-shift accommodation, including a former stable that is prone to flash flooding, and a ramshackle lean-to under a railway arch that shakes when trains thunder by, redispersing thick layers of dust and dirt.

The museum used to face directly onto the Berlin Wall, but it now has a fine view across the Spree to the spanking-new central station and the chancellery of united Germany. It belongs to the Charité, the city's historic medical school, which was reconstituted five years ago in a painful merger between the medical faculties of East Berlin's Humboldt University — the original Charité — and West Berlin's post-war Free University.

The ascendancy of the Charité over the western faculty is in part thanks to its long — if interrupted — reputation, having nurtured such scientific giants as Rudolf Virchow, the nineteenth-century pathologist who conceived the cellular basis of disease. Indeed, in the nineteenth and early twentieth centuries, Berlin was a hot-bed of modern science in all disciplines, with the likes of Hermann von Helmholtz, Carl von Siemens and Emil Fischer pushing all possible frontiers. This has bequeathed Berlin many important collections of instruments, specimens and other historic scientific items.

When the Charité was reconstituted it had a mandate to establish scientific excellence — which it has in good part done — with a shrinking budget. But what of its heritage? It has done a marvellous job in restoring the lovely old building housing the Medical Historical Museum, which now has a fine display of the cream of its collection — not to mention an extraordinary meeting room called 'The Ruin', comprising an area bombed out in the Second World War, now simply glassed around. Unfortunately this big investment is not enough. Its archived collections need a safe home. As do the collections of old DDR instruments acquired after the Wall fell, which share the same primitive lodgings. They are invaluable for historians of twentieth-century technology.

To balance its books, the Charité is selling its off-campus buildings, many of historical as well as financial value. One of the most attractive is the 1877 building on the Dorotheenstrasse, with its beautiful glass-roofed anatomical lecture theatre, where pioneering microbiologist Robert Koch announced his discovery of the tuberculosis bacillus. There are additional important collections in some of these buildings, including the physiological instruments that are this week's Hidden Treasure (see page 31). These also need new permanent homes.

To attract the best scientists, the Charité must be able to offer them generous lab space — contemporary science has, of course, to be the top priority. But heritage must not be the default victim.

The Charité will have a new management team in the autumn that would be wise to guarantee a tiny proportion of its budget to protecting the school's illustrious past — which did, after all, give it the upper hand after the reunification of Germany. The year 2010 is the 300th anniversary of the Charité, as well as the 200th anniversary of the Humboldt University and the 100th anniversary of Koch's death. This triple celebration should be a time to remember the debt to the past. Let it not be wiped out. ■

RESEARCH HIGHLIGHTS

Nasty, brutish and short

Proc. Natl Acad. Sci. USA **105,** 8980–8984 (2008)

The Madagascan chameleon *Furcifer labordi* has an annual life cycle, and spends most of its short life in the egg.

Kristopher Karsten of Oklahoma State University in Stillwater and his colleagues monitored individuals from the time of hatching until death during several cycles. They found that they hatch in November; grow at astonishing rates; reach maturity by January; battle fiercely over mates, breed and lay their eggs by February; and then promptly drop dead. For the next nine months, the entire species is represented by eggs.

This is the shortest lifespan ever recorded for a four-legged vertebrate animal.

C. J. RAXWORTHY

NEUROSCIENCE

Predicting psychosis

J. Neurosci. **28,** 6295–6303 (2008)

Scientists have found a way of predicting how an individual will respond to the party drug ketamine — and it might help us understand why symptoms of schizophrenia vary so much between individuals.

Ketamine mimics many symptoms of schizophrenia. Paul Fletcher at the University of Cambridge, UK, and his colleagues scanned the brains of 15 healthy volunteers while they performed various cognitive tasks that require skills that are often disrupted in schizophrenia, such as verbal processing and working memory.

They found, among other things, that those with higher activity in frontal, thalamic and caudate regions of the brain during a working memory task tended to become apathetic and withdrawn on ketamine. Meanwhile, those with exaggerated frontal and temporal activation responses during verbal tasks experienced disordered thoughts and abnormal auditory perception.

GENETICS

The genetics of anarchy

Genetics doi: 10.1534/genetics.108.087270 (2008)

A study of honeybee 'anarchy' has uncovered several regions of the genome that influence cheating behaviour.

Honeybee (*Apis mellifera*; pictured right) queens emit a pheromone to 'switch off' the ovaries of female worker bees, but some individuals are more sensitive to the pheromone than others. Those who fail to respond are branded anarchists because they disrupt the social order of the hive.

Peter Oxley of the University of Sydney, Australia, and his colleagues tracked down regions of the genome that have a role in ovary activation. They found four such regions that together account for 25% of the variation in this trait observed in the population of honeybees they studied.

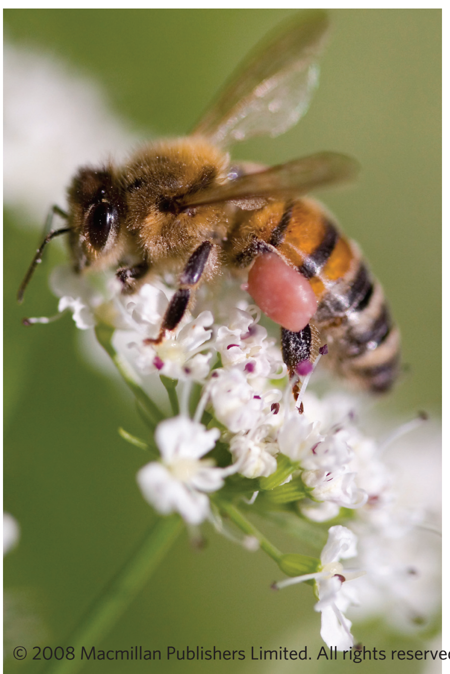
CHEMISTRY

Flipping brilliant

Organic Lett. doi: 10.1021/ol801135g (2008)

A super-fast colour-changing chemical has been synthesized by Jiro Abe and his colleagues at Aoyama Gakuin University in Sagamihara, Japan. The molecule is a ring system containing naphthalene groups.

When the colourless version of the molecule is zapped by ultraviolet light it changes to its green-coloured form by breaking a carbon–nitrogen bond to leave a molecule in which two electrons are left delocalized in their naphthalene rings as radicals. This change takes a fraction of a second. When the light is turned off the molecule can quickly flip back to its colourless version. This light-induced colour change can happen whether the molecule is a solid or in a solution. Photochromic materials such as this are used in light-sensitive lenses and data-storage devices.



L. DALTON/NHPA

CHEMICAL BIOLOGY

Anti-Alzheimer's agent

Nature Chem. Biol. doi:10.1038/nchembio.96 (2008)

Scientists have designed an enzyme inhibitor that seems to prevent the tangling of a brain protein that is linked with the onset of disorders such as Alzheimer's disease.

The inhibitor, named thiamet-G, acts by stopping the removal of sugar groups from specific sites on a protein called tau. It thereby blocks the attachment of phosphate groups thought to lead to the characteristic tangling.

The team, led by David Vocadlo at Simon Fraser University in British Columbia, Canada, tested the enzyme inhibitor in healthy rats. They found thiamet-G to be the first such inhibitor that can be delivered to the brain through the bloodstream. Besides providing a means to investigate how tau proteins form clumps, the inhibitor may have potential as a therapeutic agent.

MATERIALS SCIENCE

The heart of glass

Nature Mater. **7,** 556–561 (2008)

A glass is caught somewhere between a liquid and a crystalline solid — its atoms move, but they do so very slowly. Theorists predicted that was because the atoms arranged

themselves in localized structures that prevented both rapid flow and crystallization.

Paddy Royall of the University of Bristol, UK, Hajime Tanaka at the University of Tokyo and their colleagues have created a model of glass that helps verify these predictions. The team suspended microscopic beads in a polymer that caused them to attract one another. They watched as beads clumped into larger structures, creating a glass-like gel. The work confirms the earlier theory about localized structures and will improve the understanding of glasses, the authors write.

GENETICS

Sex and the cortex

PLoS Genet. **4,** e1000100 (2008)

How male and female brains differ is debated around the water cooler as much as the lab bench. Working at the latter, Elena Jazin at Uppsala University in Sweden and her colleagues looked for differences in gene-expression patterns in the cortex, which is associated with higher brain functions such as cognition. The team found that some human sex-specific gene-expression patterns are mirrored in the brains of other primates — macaques (*Macaca fascicularis*) and marmosets (*Callithrix jacchus*) — and that the sequence of these genes is more conserved than that of a control set of genes.

The fact that these differences are conserved across species suggests that evolution has deemed them worthy of preservation and that they may underlie some differences between the sexes, the authors say. The nature of those differences, however, remains water-cooler fodder.

ECOLOGY

Drought and the lion

PLoS ONE **3,** e2545 (2008)

Extreme weather can cause mass die-offs in the animal kingdom by altering host–pathogen relationships, according to researchers led by Craig Packer of the University of Minnesota in St Paul.

They found that high lion mortality in Tanzania in 1994 and 2001 seemed to be linked to severe drought followed by heavy rain. Blood work on lions suggested the following explanation: the drought-starved buffalo had heavy tick infestations after the rains, often killing them and providing carcasses for lions to scavenge. A coincident epidemic of normally non-fatal canine distemper virus suppressed the lions' immune systems, allowing the also normally harmless tick-borne blood parasite *Babesia* to reach fatal levels.

ACOUSTICS

Fiddling the numbers

PLoS One **3,** e2554 (2008)

Subtle shifts in density that occur within individual pieces of wood might help to explain why violins made in eighteenth-century Cremona, Italy, sound so special.

Berend Stoel, of Leiden University Medical Center in the Netherlands, and Terry Borman, a violin maker based in Fayetteville, Arkansas, measured the density of five classical violins — including two made by Antonio Stradivari (such as that pictured below) — and eight modern instruments, using computed tomography. The difference in density between spring- and summer-growth spruce and maple was significantly smaller in the classical instruments than in the modern ones. Stoel and Borman suggest that these variations in density may influence the wood's acoustic properties by affecting its stiffness.

L. PITARAKIS/AP

NANOTECHNOLOGY

Electron windmills

Phys. Rev. Lett. **100,** 256802 (2008)

Carbon nanotubes can be sent spinning by passing an electrical current through them, Steven Bailey and his colleagues at Lancaster University, UK, say.

Their calculations show that electrons passing through a nanotube with a typical 'chiral' structure, in which helical ribbons of carbon hexagons wind around the tube axis, will pick up angular momentum on the way and create a twisting torque on the tube. For a tube nested inside the sleeve of a wider one, this force should be big enough to overcome frictional resistance and drive rotation, auguring well for nanoscale drills and motors.

JOURNAL CLUB

Peter Csermely
Semmelweis University, Budapest, Hungary

A network scientist highlights active sites of enzymes, cells, brains and society.

For proteins, chemical binding is a tricky business. Special signals must be sent across a sea of water molecules to the desired partner, and complex mutual structural adjustments (a fluctuation fit) must be completed before each successful binding event.

I have long taught that a protein at its lowest-energy conformation still has regions of higher energy. But I've always been intrigued: how is the extra energy of the active sites preserved? And why do we need such big enzymes when their active sites occupy only a tiny region?

Piazza and Sanejouand found part of the answer by identifying special energy-preserving segments of proteins (F. Piazza and Y.-H. Sanejouand *Phys. Biol.* **5,** 026001; 2008). Taking into account the effect of the surrounding water, they modelled proteins with a computer program that arranges oscillating elements in the same pattern as amino acids in real proteins. In most of these proteins, they identified a few easily excitable segments that collected and harboured long-lived, localized vibrations. An analysis of 833 enzymes showed that these segments co-occur with the catalytic active sites; are located on the stiffest parts of the proteins; and have many connections but are surrounded by a less well-connected environment.

The generality of many network properties prompts me to ask: can we find 'active sites' of cells, brains, ecosystems and societies? Piazza and Sanejouand's segments correspond to Ronald Burt's "structural holes" in social networks — whereby areas of greatest economic potential are areas of low connectedness, where brokers can make new connections. Indeed, not only amino acids, but people may also act as brokers, mediators and catalysts. It may be worthwhile to think about creative, broker proteins as drug targets. One could even imagine creative sets of neurons.

Discuss this paper at **http://blogs.nature.com/nature/journalclub**

NEWS

Neuroscientist: my data published without authorization are 'misleading'

The director of a top laboratory in Germany has charged that two of his former research students took data from his laboratory without his permission and published scientifically incorrect interpretations of them against his advice.

Neuroscientist Nikos Logothetis (pictured), of the Max Planck Institute for Biological Cybernetics in Tübingen, further claims that the journal involved, *Human Brain Mapping*, acted incorrectly by publishing the paper after he told them the data were inappropriate. He says the journal has denied him the right to a timely reply.

One of the two editors-in-chief of *Human Brain Mapping*, Peter Fox of the University of Texas Health Science Center in San Antonio, told *Nature* that the paper was correctly refereed, but declined to add details.

"The journal used the Max Planck Society to excuse their own mismanagement of the case."

Logothetis is furious about the publication of data, which he believes will mislead the field, and about the fact that the authors of the paper allege that he tried to stop them publishing the data for personal reasons.

The affair began in the spring, when Amir Shmuel, who worked in Logothetis's laboratories from 2002 to 2007 and is now at the Montreal Neurological Institute of McGill University in Canada, asked Logothetis for permission to use data generated there.

Although he agreed at first, Logothetis withdrew his permission when he realized that the data — from functional magnetic resonance imaging studies on monkey brains — were being used to support a theory about spontaneous brain activity. The data had been collected when monkeys were looking at a grey but flickering LED screen. "The protocol was just inappropriate for analysis of spontaneous brain activity," says Logothetis.

Several months later, he says, he was surprised to receive an e-mail from Shmuel containing a complete paper using the same data, co-authored with another former research student, David Leopold, who worked in the labs between 1992 and 2003, where he collected some of the data himself. Leopold is now at the National Institute of Mental Health in Bethesda, Maryland. Shmuel invited Logothetis to join as third author, telling him that the paper had already been accepted for publication and would appear online in a few days. It had been accepted six weeks earlier.

Matters escalated. "I told him that the data were not publishable," says Logothetis, who also wrote to Fox proposing that the paper should not be published. But Leopold then wrote directly to the Max Planck Society (MPS), which runs 80 research institutes in Germany, claiming that Logothetis was trying to prevent him and Shmuel from publishing data for personal reasons.

After consultation with Logothetis, MPS vice-president Herbert Jäckle wrote to the authors giving approval for the use of the data,

Turkish politics blamed for board block

A prominent Turkish geologist is being denied a top spot in the nation's higher-education system because, he says, his political views are out of step with those of the current government.

Celâl Şengör (pictured), a professor at Istanbul Technical University, says that he has been blocked from joining Turkey's council of higher education (YÖK) and subjected to spurious ethics investigations because he has spoken out against the government.

"There is an atmosphere of terror," he says. "It's unbelievable what's going on."

YÖK is a 21-member council that oversees Turkey's universities. One-third of the council is nominated by the inter-university board — a group of university rectors and other academic representatives. In January, the board advanced Şengör's name for a position on YÖK. The reason for the nomination was straightforward, says Mustafa Akaydin, who heads the inter-university board. "He's a good guy, a very well-known intellectual in Turkey."

Nominations typically pass through a straightforward approval process, but Şengör says that YÖK's leadership is holding up his candidacy. This spring, he says, the leadership opened an investigation into a previously dismissed ethics charge concerning Şengör's ties to a family business. Then, just last week, Şengör says he was informed of a second investigation into his having allegedly travelled abroad without university approval. Şengör believes that he is being persecuted because he is an outspoken critic of Turkey's current, Islamist-rooted AKP ruling party. "They want to get rid of anybody who is against them," he says.

AP

NUCLEAR KNOCKDOWN Is North Korea really decommissioning? www.nature.com/news

but adding that Logothetis's scientific concerns should be taken into account, in accordance with the MPS's code of good scientific practice. Small changes were made to the paper that did not satisfy Logothetis, and it went online as planned on 8 May.

Jäckle says that the journal misrepresented his approval of the use of the data as being permission to publish. "But we only ever gave approval to use the data — the journal used the Max Planck Society to excuse their own mismanagement of the case."

Jäckle also notes that he requested that the MPS should not be listed in the paper as a funder of the project, because neither of the two authors had been directly funded by the society. But the request was ignored.

Logothetis says the paper does not give sufficient information to have allowed referees to understand the source of the data. He adds that *Human Brain Mapping* has not guaranteed him the opportunity to publish a response with his own interpretation of the data. He says the paper could mislead the field, for example, with its claims to see waves of activity in the cortex that Logothetis's analysis does not support.

Fox, who refused to retract the paper, says: "The editors of *Human Brain Mapping* were entirely appropriate — that's all I want to say about it."

Shmuel and Leopard issued a statement to *Nature* in which they say: "We are confident, and rigorous peer review agreed, that the data we collected are appropriate for studying spontaneous activity and the resting state in the brain. We stand by the conclusions we made in our paper." ■

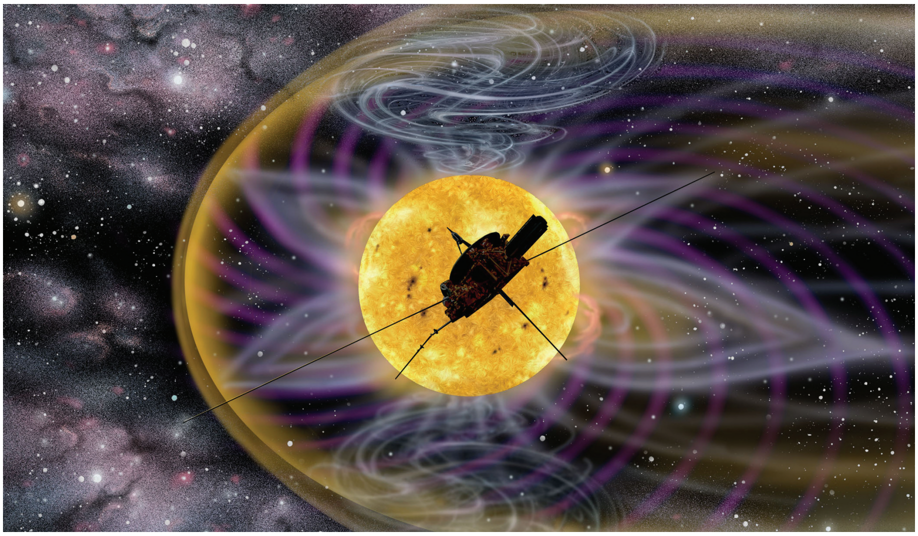
Alison Abbott

Şengör has also spoken out publicly against YÖK's newly appointed president, Yusuf Ziya Özcan, a US-trained sociologist. In the Turkish press he has stated that Özcan is not qualified to be an assistant professor, let alone head of Turkey's higher-education council. "I simply said that this man is incompetent," Şengör says.

Such blunt statements have probably not endeared him to YÖK's leadership, says Şevket Ruacan, former head of Tübitak, Turkey's national science research agency. Nevertheless, Ruacan says that the inter-university board's nomination should be respected. "The president has no option but to appoint Celâl," he says.

Özcan declined *Nature*'s request for an interview. ■

Geoff Brumfiel



D. HARDY/ESA

Closer than ever to the Sun

As the Voyager missions carry the hopes and dreams of Earthlings to the edge of the Solar System (see page 24), work is beginning on a major new mission to explore the superstar at its heart.

But scientists must first say goodbye to its predecessor Ulysses, the longest-serving spacecraft, which has swooped past the Sun's poles three times on its unique orbit out of the elliptic plane of the Solar System. Ulysses has been a joint project between the European Space Agency, which built the spacecraft, and NASA, which provided the launch rocket and a plutonium-238-powered energy source. Its main goal was to map the heliosphere, the Sun's bubble of influence created by a stream of charged particles called the solar wind — which, for instance, 'blows' a comet's ion tail so that it always points away from the Sun. Ulysses passed through several comet-tails and measured the solar wind at periodic lows and highs, revealing that the wind is quite variable in time and space. Ulysses also discovered that the Sun's magnetic field has a strange asymmetry between its north and south poles.

But after nearly 18 years, the mission's odyssey is over. The fuel line could freeze shut in days, according to Ed Massey, NASA project manager at the Jet Propulsion Laboratory in Pasadena, California. There is no Ithaca for this Ulysses — once the fuel line freezes, engineers won't be able to orient the spacecraft's radio antennas towards Earth. It will continue to loop blindly around the Sun until at least 435 years from now, when a close pass of Jupiter could alter its course.

While Ulysses gasps its last, another NASA mission, Solar Probe Plus, is in initial development at the Applied Physics Laboratory in Laurel, Maryland. Solar Probe Plus would get closer to the Sun than any mission before in order to address two mysteries: why is the corona — the wispy, million-degree part of the Sun's atmosphere that can be seen during eclipses — so much hotter than the 6,000-degree surface of the Sun? And how, in this region, is the solar wind accelerated to supersonic speeds?

Protected by a carbon-foam shield 2.7 metres in diameter, Solar Probe Plus would tiptoe to the edge of the corona, passing at its closest within 6 million kilometres of the Sun's surface — about 10 times closer than Mercury.

For a solar mission, Ulysses never really got all that close. In fact, it was at its closest when it launched from Earth. And so it measured the effect of the Sun on the solar wind, rather than getting to the roots of its cause. "It's like trying to understand hurricanes without incorporating oceans in the model," says Andrew Dantzler, the project manager for Solar Probe Plus.

The US Congress specifically funded $14 million in 2008 towards mission-technology development, but NASA did not request any money for it in 2009. The agency plans to ramp up mission funding, spending $220 million by 2013 towards a tentative 2015 launch. But budget outlooks are still well short of the spacecraft's proposed $900-million overall cost. ■

Eric Hand

Reincarnation can save Schrödinger's cat

It's one of the most perplexing questions in physics: how does the seemingly exotic behaviour of tiny particles in the quantum realm collapse to create the classical reality observable in matter that is at least a molecule big? Now, an experiment further muddies the distinction between the two realms by demonstrating that it is possible to halt the transition from the quantum to the classical in its tracks — and reverse it. The achievement could provide quantum computing with a crucial capability (See 'A fix for quantum computers').

According to the 'Copenhagen interpretation' of quantum mechanics, arrived at in 1927, observing a quantum object disturbs its state, causing it to flit from quantum to classical realities. Before they are measured, atoms and subatomic particles do not have fixed properties, they exist in a 'superposition' in which they have many mutually contradictory properties.

The notion is exemplified by the paradox of Schrödinger's cat, a thought experiment in which a cat is locked in a box with a vial of poisonous gas that would be broken if a quantum particle was in one state, and remain intact if the particle was in another. While the box is closed, the particle exists in a superposition of both states simultaneously, so the poison must also simultaneously be both released and contained, and, in turn, the cat must be both alive and dead. When the box is opened, the quantum superposition collapses, and the cat is either killed or saved, in an instant.

Now, Nadav Katz at the University of California, Santa Barbara, and his colleagues have performed an experiment in which they pull a quantum state back from the brink of collapse, 'uncollapsing' it and returning it to its unobserved state. Effectively, they have peeked at Schrodinger's cat in its box, but saved it from near-certain death (N. Katz *et al.* http://arxiv.org/abs/0806.3547).

To physicists raised on the textbook Copenhagen interpretation, any notion of uncollapsing a quantum state seems "astonishing", says Markus Büttiker, a quantum physicist at the University of Geneva in Switzerland. "On opening the box, Schrödinger's cat is either dead or alive — there is no in between."

However, a more recent interpretation of quantum mechanics, 'decoherence theory', suggests that collapse does not occur instantaneously. Instead it plays out gradually as the quantum system slowly interacts with its environment (see *Nature* **453**, 22–25; 2008). In 2006, Alexander Korotkov of the University of California, Riverside, and Andrew Jordan, of the University of Rochester in New York, proposed that this may leave open a time period in which experimenters could intervene to halt the collapse (A. N. Korotkov & A. N. Jordan *Phys. Rev. Lett.* **97**, 166805; 2006). They provided blueprints for an experiment to test the idea, which Katz, Korotkov and their colleagues have now done.

Living dead

In place of Schrödinger's cat, Katz and his colleagues created a 'phase qubit', often used in quantum computing experiments, involving a superconducting circuit that is broken by an insulating junction. The qubit is made up of fluctuations in the average current flowing around this loop and is characterized by its 'phase', the degree that the current gets knocked out of step as it crosses the junction.

The qubit can have two different energies, high and low, and the team prepared the qubit so that it took on a superposition of both energy values simultaneously — an experimental equivalent of Schrödinger's cat being simultaneously alive and dead. Any attempt to measure the qubit's energy directly would permanently collapse the state into one of these two energy values — effectively opening Schrödinger's box. The trick is to avoid this total collapse by sneakily getting an indirect hint of the qubit's energy, explains Katz.

The team turned to a quantum property known as 'tunnelling', whereby quantum particles can overcome energy barriers when they seem to have too little energy to do so.

In this case, the team changed the average current running through the circuit to control the 'height' of a 'barrier' that would make it harder for the qubit to change phase. The barrier was strong enough to prevent a low-energy qubit from changing to a new phase, but a high-energy qubit state could be tempted to make

A fix for quantum computers

Quantum computers promise to outperform their classical counterparts — but they are tough to construct. However, the ability to 'uncollapse' quantum states demonstrated by Nadav Katz at the University of California, Santa Barbara, and his colleagues could now give those attempting to build such computers a boost.

Conventional computers process information in 'bits' that take on values of 0 or 1. Quantum computers, by contrast, use 'qubits' that can also exist in 'superpositions' of both states simultaneously. This means that a quantum computer can carry out multiple calculations in parallel. But so far, only a relatively small number of qubits have been combined to perform calculations in the lab.

One of the biggest problems plaguing those hoping to build large-scale quantum computers is that qubits are extremely fragile. Even a tiny interference from the outside world can knock a qubit out of whack, collapsing its quantum state and leading to loss of stored information.

The demonstration that it is possible to rescue a collapsing qubit, 'uncollapse' it and return it to its original quantum state could one day be used to rectify errors before information is lost.

Robin Blume-Kohout, a quantum theorist at the Perimeter Institute for Theoretical Physics in Ontario, Canada, agrees. "This might be a very useful error-correction technique for quantum information processing down the road." **Z. M.**

the transition. So, looking at whether the qubit made the transition — creating a telltale burst of magnetic energy — can reveal something about the energy state of the qubit.

The least exciting result is if the qubit succeeds, says Katz. This means that it definitely fully collapsed into its high-energy state and tunnelled. "That's game over," says Katz. It's the equivalent of fully opening the box, seeing that the cat is definitely alive, and releasing it.

Time travel

It's when the qubit does not tunnel that things become more interesting. This means there is a good chance that it was in the lower level. However, crucially, you cannot know for sure, so this measurement does not completely collapse the system, says Katz. "We have peeked at the cat and then quickly shut the lid."

This 'weak' measurement does slightly disturb the system, however. If the experiment ended there, this disturbance would be enough to slowly drive the qubit towards a total collapse into the lower energy state, says Katz. His team has statistically confirmed this outcome by repeating the measurement on thousands of similarly prepared qubits and then terminating the experiment at this point. Measuring the qubit's state at this stage reveals that in the vast majority of cases, it collapses to the lower energy state. That is, on reopening the box, the cat is most likely to be found dead.

To uncollapse the state, the team had to catch the qubit before it completed this journey to complete collapse and "undo the damage that we had done", says Katz. They used a standard technique to swap the qubit's energy levels, by firing a specially prepared microwave pulse at it — turning a qubit in a high-energy level to one in a low level, and vice versa. Because of the swap, when they then repeated their weak measurement, they caused a disturbance that exactly cancelled the effects of their first measurement. The first time they peeked, they nudged the cat towards death, but the second time they nudged it towards life, leaving it back where it started, explains Katz.

Once again, the team confirmed statistically that the state had been returned to its original uncollapsed form — with a roughly equal probability of collapsing into either energy level — by performing multiple runs of the experiment. Measuring the state at the end of the experiment, they found that it was just as likely to be found in either the high- or low-energy state.

"The data are clear," says Büttiker. "This is a breakthrough experiment."

The result is a warning that our understanding of how classical reality emerges may be naive, says Vlatko Vedral, a quantum physicist at the University of Leeds, UK. "It tells us that we really can't assume that measurements create reality, because it is possible to erase the effects of a measurement and start again."

"The quantum world has become more tangible, and the nature of reality even more mysterious," adds quantum theorist Maximilian Schlosshauer of the University of Melbourne in Australia. ■

Zeeya Merali

LANE DISCIPLINE? Olympic sprints hits more controversy **www.nature.com/news**

INGRAM PUBLISHING (SUPERSTOCK)/ALAMY

PLoS stays afloat with bulk publishing

Public Library of Science (PLoS), the poster child of the open-access publishing movement, is following an *haute couture* model of science publishing — relying on bulk, cheap publishing of lower quality papers to subsidize its handful of high-quality flagship journals.

Since its launch in 2002, PLoS has been kept afloat financially by some US$17.3 million in philanthropic grants. An analysis by *Nature* of the company's accounts shows that PLoS still relies heavily on charity funding, and falls far short of its stated goal of quickly breaking even through its business model of charging authors a fee to publish in its journals. In the past financial year, ending 30 September 2007, its $6.68-million spending outstripped its revenue of $2.86 million, according to the publicly available accounts.

But its financial future is looking brighter thanks to a cash cow in the form of *PLoS One*, an online database that PLoS launched in December 2006. *PLoS One* uses a system of 'light' peer-review to publish any article considered methodologically sound. In its first full year of operation in 2007, *PLoS One* published 1,230 articles, which would have generated an estimated $1.54 million in author fees, around half of PLoS's total income that year. By comparison, the 321 articles published in *PLoS Biology* in 2007 brought in less than half this amount.

"PLoS trumpeted its business model as being better than everyone else's."

From the outset, the company consciously decided to subsidize its top-tier titles by publishing second-tier community journals with high acceptance rates that would be cheaper to produce. In addition to its premium titles — *PLoS Biology* and *PLoS Medicine*, which have low acceptance rates and high overheads, and charge author fees of $2,750 — it has launched four lower-cost journals that are run by volunteer academic editorial teams rather than in-house staff. At $2,100, the fees to authors for publishing in these journals — *PLoS Computational Biology*, *PLoS Genetics*, *PLoS Pathogens* and *PLoS Neglected Tropical Diseases* — are almost as much as those for its top journals.

But this combination has generated much less income than PLoS envisaged. "It's very clear that the reality is very far from PLoS's original plan," says John Ewing, executive director of the American Mathematical Society. In particular, PLoS's initial projections overestimated by an order of magnitude the number of articles that both its top- and second-tier journals would attract, tax-record data obtained by *Nature* reveal. PLoS also grossly underestimated the cost of producing its top-tier journals.

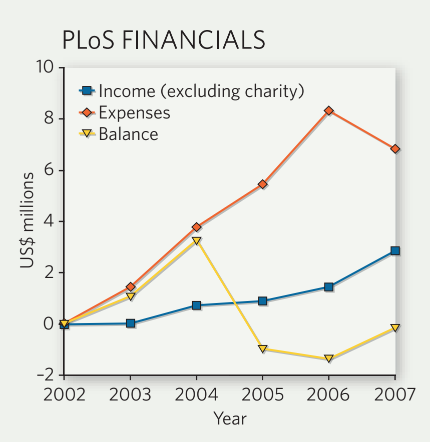
"If the original model was to be self-sustaining through author fees, it seems that PLoS is not even half-way there," says Bernard Rous, deputy director of publications of the Association for Computing Machinery, the world's largest educational and scientific computing society. Nevertheless, Rous endorses PLoS's strategy of tapping multiple revenue sources and cross-subsidizing to allow open access to all its titles.

Faced with an increasingly dire financial situation, PLoS hiked all its author fees in 2006, from $1,500 to as much as $2,500 for its top two journals, and has since increased them further. This has helped to boost revenue, but the effect of the increase has been minor compared with the new cash flowing from the flood of articles appearing in *PLoS One*. Since it launched, *PLoS One* has published 2,526 articles — more than *PLoS Biology* has published in its five years of existence, and that growth shows no signs of abating. *PLoS One* has published 1,158 papers since the beginning of this year, which is almost as many as it published during the whole of 2007. Another factor is that it costs authors only $1,250 to publish in *PLoS One*.

"It's fair to say that the community-run journals, including *PLoS One*, are contributing very well to our overall financial picture," says Peter Jerram, chief executive of PLoS, adding: "PLoS is on track to be self-sustaining within two years. In the interim some philanthropic support will be needed."

"It's tremendous to see how the progress that PLoS has made over the years has been mirrored by the great strides taken towards open-access to research more broadly," Jerram says.

Papers submitted to *PLoS One* are sent to a member of its editorial board of around 500 researchers,



who may opt to review it themselves or send it to their choice of referee. But referees only check for serious methodological flaws, and not the importance of the result. The board members who *Nature* spoke to were generally positive about the overall quality of the papers they had received to review and the referees reports they solicited.

"There's so much in *PLoS One* that it is difficult to judge the overall quality and, simply because of this volume, it's going to be considered a dumping ground, justified or not, says John Hawley, executive director of the free-access *Journal of Clinical Investigation*. "But nonetheless, it introduces a sub-standard journal to their mix."

Hawley says it's unsurprising that the PLoS business model has come under scrutiny. "PLoS trumpeted its business model as being better than everyone else's, as being 'the one'." The tack taken by BioMed Central, a London-based open-access company, of publishing mostly lower-impact journals in a wide variety of disciplines "is probably closer to what works in open access", he says.

BioMed Central has an estimated annual revenue of around £10 million ($20 million). It is already "pleasantly profitable", according to a science-publishing consultant who asked to remain anonymous. "BioMed Central knows well that much of the journal middle order is more profitable than the great brands because of the lower editorial costs and the cheaper marketing costs for bundles of journals. I suspect that *PLoS One* is a result of learning the same lesson," he adds.

BioMed Central is now up for sale, which will be a "fascinating first market test of what people will pay for an open-access company". ■

Declan Butler

SIDELINES

ON THE RECORD

“Iceland does not subscribe to the principle of survival of the cutest.”

Stefan Asmundsson, Iceland’s representative to the International Whaling Commission, defends the nation’s position on “sustainable utilization” of whales.

SHOWBIZ NEWS

Knight guider

It’s now possible to get a Global Positioning System in-car map reader that talks to the driver in the voice of KITT, the intelligent car from the classic 1980s TV show *Knight Rider*. Sadly, this won’t enable a car to drive itself — even if a Hoff-alike shouts: “I need ya buddy.”

THE KOBAL COLLECTION/UNIVERSAL TV

NUMBER CRUNCH

122,500 is the number of species identified by the Census of Marine Life so far.

56,400 is the number of ‘double identities’ that have been eliminated in the process, ensuring that each species is catalogued only once.

56 is the record number of aliases for a single species, accrued by the breadcrumb sponge (*Halichondria panicea*) since 1766.

230,000 is the number of marine species thought to be known to science. It is hoped that they will all be on the census by 2010.

ZOO NEWS

Fish out of water

Lucky pet goldfish are getting a penthouse view in their new ‘supmarine’. The inverted submarine, an upside-down, water-filled glass bowl was invented by a Dutchman for his son. It sits above the water level, so fish can swim up into it and look out.

Sources: Reuters, PC Magazine, Census of Marine Life, Wired

Oil and food costs may bump climate at G8 talks

The world’s most pressing problems will be the focus of G8 talks next week, when leaders from rich nations and major emerging economies meet in Toyako on the Japanese island of Hokkaido.

Formed during the 1970s oil crises, the G8 (Group of Eight) is an exclusive but informal bloc of nations — comprising the world’s largest developed economies — that meets annually to consider how to tackle problems of global importance.

More than 30 years on, oil prices are again high on the agenda for the G8 gathering, which has climate change as its main focus. It is also expected to cover global food shortages, nuclear non-proliferation, African development and peace building.

In addition to the usual G8 heads, European environment commissioner Stavros Dimas will attend this year’s summit. The meeting will expand to include a further eight industrialized nations on the final day to facilitate the Major Economies Meeting on Energy Security and Climate Change, an initiative floated last year by US President George W. Bush to address such issues.

After the political breakthrough of the 2007 summit in Heiligendamm, Germany, where delegates agreed to seriously consider slashing emissions to half of 1990 levels by 2050, leaders meeting in Japan are under pressure to seal a deal on long-term targets for reducing greenhouse gases.

“A year after saying they would seriously consider cutting emissions 50% by 2050, it is now incumbent upon them to agree to it,” says Elliot Diringer, director of international strategies at the Pew Center on Global Climate Change in Arlington, Virginia.

The major economies failed to agree on a draft G8 accord on emissions targets last month in Seoul, South Korea, causing speculation that little progress will be made on the issue. Still, Diringer says: “I wouldn’t rule it out, [though] it would require a shift on the part of the Bush administration.”

But scientists and environmentalists have warned that shorter-term targets are needed to avert dangerous climate change. United Nations climate chief Yvo de Boer has urged G8 leaders to set targets for 2020 during the summit in Japan.

Whether emerging economies such as China and India should be subject to emissions limits remains a serious sticking point, however, and one that is unlikely to be resolved during two days of talks. Some say it might be more realistic to expect the forum to agree on more immediate actions that could provide tangible results. On 10 June, the science academies of the ‘G8+5’ countries — the G8 plus China, India, Brazil, Mexico and South Africa — released a statement calling on the summit to assist the transition to a low-carbon economy by funding demonstration projects for carbon capture and storage.

Y. TSUNO/AFP/GETTY IMAGES

The pressure is on for the G8 to come up with global solutions.

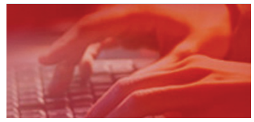
“The key would be to agree to something that can make a difference. What could really make a difference is speeding up carbon capture and storage technology,” says Martin Rees, president of Britain’s Royal Society, one of the statement’s signatories.

“This is an opportunity where the G8 countries with the United States on board could make a declaration that is feasible and financially modest given the scale of the problem,” he adds.

But with soaring oil and food prices shaking economies around the world, commentators speculate that global warming could be bumped down the agenda at next week’s talks. G8 leaders are expected to use the meeting to call on nations to consider releasing stockpiles of food and to restrain from curbing exports. ■

Olive Heffernan

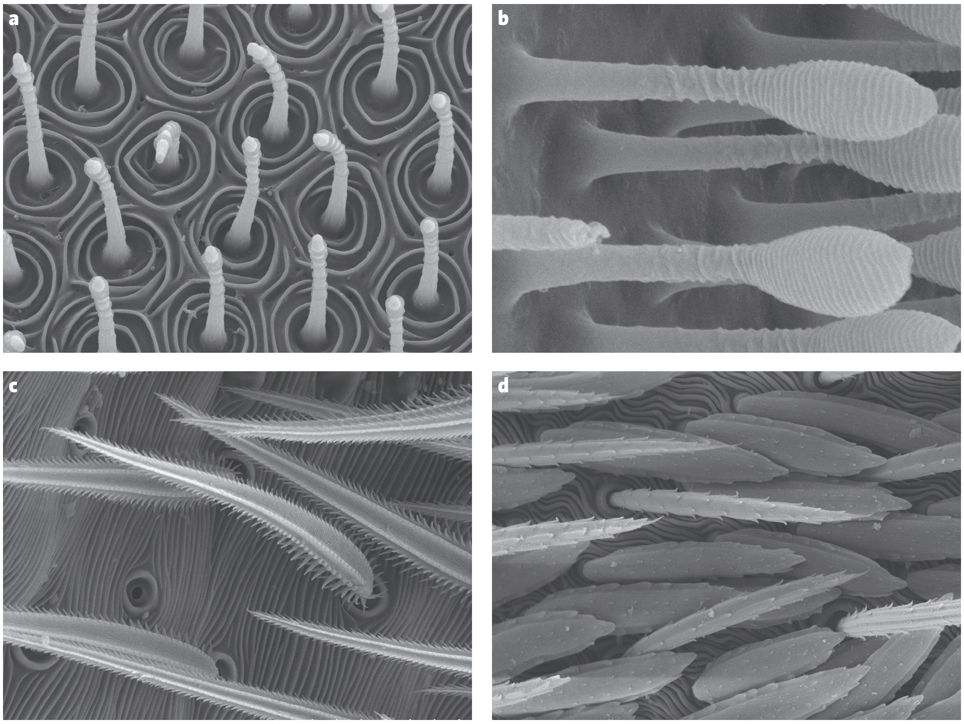
For the latest from the G8 meeting, visit http://blogs.nature.com/climatefeedback.



THE GREAT BEYOND
Nature's news blog.
http://blogs.nature.com/news/blog/

SNAPSHOT

Incy Wincy spider



A field of spigots is seen in this electron microscope close-up (a) of the *Thaida peculiaris* spider's silk-spinning organ, or cribellum. The spigots, seen even closer in an image of the *Filistata insidiatrix* spider (b), extrude liquid silk through openings at the bulbous end. The pressure of the extrusion and manipulation of the silk by the spider's legs transforms liquid silk into a flexible thread. Spider hairs, or setae, on the abdomen (c) can contain sensory neurons and irritating chemicals for attacking enemies. The scales (d), made of chitin, are responsible for the brilliant colour and iridescence of some spiders.

These are four of more than 32,000 images from the Phylogeny of Spiders project, part of the international, US-funded Tree of Life initiative to create an evolutionary map of all plants and animals using genomic and physical data. ■

Rex Dalton

NATURE|Vol 454|3 July 2008

Entomologists jailed for illegal specimen hunt

Indian scientists are being asked to help facilitate the release of two Czech entomologists who have been detained in an Indian jail since 22 June.

Petr Švácha, of the Institute of Entomology at the Biology Centre of the Academy of Sciences of the Czech Republic in České Budějovice, and his colleague Emil Kučera were arrested for collecting beetles and butterflies without permission from a national park in West Bengal, in violation of the Indian Biodiversity Act. Arguments are set to be heard on the case on 7 July in Darjeeling.

Although the Indian media has accused them of collecting insects for the Chinese medicine market, the Czechs claim that they were unaware that collecting insects was illegal in India. "These people are sincere, genuine entomologists, and the specimens that they have collected are of no commercial value," says Max Barclay, a senior curator at the Natural History Museum in London, who is calling for their release.

Jan Sula, director of the Czech institute, told Indian authorities that "we believe the incident resulted from misunderstanding as both men did not realize that they had entered the national park".

Leading invertebrate scientists in India are planning to submit a joint petition to the Indian prime minister and the chief minister of West Bengal to free the researchers.

Muslim nuclear physicist sues over freedom of rights

An Egyptian-born nuclear physicist is suing the US Department of Energy over the revocation of his security clearance.

Moniem El-Ganayni worked from 1990 as a scientist at the Bettis Laboratory, a contract lab for the energy department near Pittsburgh, Pennsylvania, that works on nuclear propulsion. El-Ganayni, a US citizen since 1988, is an active member of the local Islamic community. He has been an outspoken critic of the Iraq war and the US government's treatment of Muslims, and was particularly critical of government efforts to recruit informants, according to his lawyer Keith Whitson, a partner at Schnader Harrison Segal and Lewis in Pittsburgh.

In May, the energy department revoked El-Ganayni's security clearance, in part because he was suspected of "conflicting allegiances". Shortly thereafter, he was fired from the laboratory. El-Ganayni's suit, filed with the help of the American Civil Liberties Union, claims that the Department of Energy has violated his rights of free speech and freedom from discrimination.

K. SRAKOCIC/AP

Moniem El-Ganayni is an outspoken critic of the US government.

Linnean Society celebrates seminal evolution papers

This week 150 years ago, papers by British naturalists Charles Darwin and Alfred Russel Wallace were read at the Linnean Society of London, proposing a radical theory that would remove humanity from the centre of creation: the evolution of all life by a process of natural selection.

On Tuesday, society members recreated the 1 July 1858 reading at the Royal Academy on Piccadilly. Darwin went on to develop his idea, publishing his seminal work *On the Origin of Species by Means of Natural Selection* in 1859, whereas Wallace slipped into relative obscurity (see *Nature* 453, 1188–1190; 2008).

Final decision expected on European stem-cell patent

The highest authority of the European Patent Office (EPO), the Enlarged Board of Appeal, will make a final decision within the next few weeks about whether human embryonic stem cells can be patented in Europe.

A public hearing at the EPO's headquarters in Munich, Germany, last week considered a 1996 patent application by the Wisconsin Alumni Research Foundation (WARF) that covered methods of cultivating primate embryonic stem cells, as well as the cell lines themselves. WARF appealed after its application was rejected in 2004. The EPO disallows any patenting on "uses of human embryos for industrial or commercial purposes".

The board's decision will affect not only the Wisconsin patent, but at least 41 similar pending patents in Europe.

Spain awards apes legal rights

The Spanish parliament's environment committee last week approved resolutions for chimpanzees, gorillas and orangutans to gain some statutory rights currently applicable only to humans. It is thought to be the first time a national legislature has taken such action.

The resolutions, which passed with cross-party support and are expected to be approved as laws by the full parliament within a year, are based on the Great Ape Project, a framework designed by scientists and philosophers who believe that humans' closest biological relatives deserve the right to life, liberty and protection from torture.

The laws will ban potentially harmful research, ape trading, profiting from apes, and using apes in performances. Zoos could still legally hold apes, but living conditions must be "optimal".

I. ARNDT/NATUREPL.COM

Pay-off agreed for expert fired after anthrax attacks

A biological-weapons researcher will receive US$5.85 million to drop his civil case against the US Department of Justice, which publicly named him in association with a series of anthrax attacks in 2001.

Steven Hatfill, a former employee of the US Army Medical Research Institute of Infectious Diseases at Fort Detrick in Frederick, Maryland, was named by then-attorney general John Ashcroft as a "person of interest" after a string of anthrax mailings to media outlets and lawmakers. Hatfill was followed, his phones were tapped and he lost his job at Louisiana State University in Baton Rouge (see *Nature* 419, 104; 2002).

Hatfill said that this had violated his privacy rights. As part of the terms, the government admits no wrongdoing in the case.

Getting it across

Scientists need a carefully crafted strategy to catch the attention of policy-makers. **David Goldston** explains.

PARTY OF ONE

Scientists often think of the US Congress as a kind of impenetrable fortress in a foreign land, nearly impossible to fathom or infiltrate. But a better analogy is something more familiar to researchers — a semi-permeable membrane. The passage of ions through such a membrane depends on concentration, temperature and pressure. The same can be said about the way Congress absorbs ideas.

This occurred to me last month when a House of Representatives committee approved a bill to promote and coordinate research on ocean acidification. A Senate committee had already approved a version of the measure, and provisions on ocean acidification were also included in the 2006 rewrite of the primary US fisheries law and in the leading climate-change bill, which the Senate killed last month. How did Congress (or at least elements of it) come to care about an issue that was barely a blip on the scientific radar screen five years ago?

The answer includes concentration (figuring out how to make the issue more potent), temperature (turning up the heat on the issue through the media) and pressure (pushing selected members of Congress and their staffs).

In this case, concentration involved at least two steps: getting the attention of the larger scientific community, and deciding what to call the phenomenon. As of 2004, a number of key papers had been published on how carbon dioxide could alter the ocean (see pages 46 and 96), but many scientists still saw oceans mainly as a carbon sink that dampened impacts on the climate. When a major international conference on the oceans and carbon was being put together for May 2004, the initial focus was on getting the world's seas to sequester more carbon*.

But scientists working on acidification, such as Ken Caldeira of the Carnegie Institution of Washington in Stanford, California, convinced the organizers to broaden their focus. In the end, 'The Ocean in a High CO_2 World' meeting in Paris, sponsored by the United Nations Educational, Scientific and Cultural Organization (UNESCO), became a turning point in expanding awareness among scientists about 'acidification'. It was also where researchers agreed on using that term, replacing more technical language about saturation, or reduced alkalinity, or levels of aragonite. Acidification was easy to comprehend, sounded alarming, and drove home the idea that carbon dioxide was a pollutant. The renaming was a critical step, Caldeira told me in an interview.

The public, though, remained largely unaware of the problem. There were a few articles in the United States on the UNESCO meeting, but media coverage increased when Britain's Royal Society published a report sounding the alarm on acidification in June 2005. An article in *Scientific American* in March 2006, by Scott Doney of the Woods Hole Oceanographic Institution in Massachusetts, garnered more attention, and the issue entered the limelight when *The New Yorker* ran a lengthy piece by journalist Elizabeth Kolbert in November 2006. These publications reached critical, influential audiences even if they did not make acidification a household word.

Even before the piece in *The New Yorker*, press coverage began to make a difference. Spurred by an article (he doesn't remember which one), Representative Jay Inslee (Democrat, Washington) set up a dinner briefing for a handful of interested colleagues in May 2006. The meeting was arranged through COMPASS (the Communication Partnership for Science and the Sea), a foundation-funded scientific group established in 1999 to link researchers with policy-makers and the media. COMPASS, in effect, helps to ferry scientists across the semi-permeable membrane, makes sure their presentations are properly structured, and ensures that they bind with the right sites within Congress.

The dinner briefing was given by Caldeira and Joanie Kleypas of the National Center for Atmospheric Research, who followed up with a briefing for congressional staff the next day. (Briefing slides at www.compassonline.org.) COMPASS helped to ensure that the presentation would be understood by policy-makers, Caldeira says. Scientists "don't understand what's scientific jargon", he notes. "Even terms such as 'water column' can be confusing." Since that event, COMPASS has taken the lead in setting up more meetings on ocean acidification between scientists and key congressional members and, especially, staff.

As with most issues, what is needed for action, especially at first, is a small group of devoted players on the relevant congressional committees, rather than broad support. In this instance, the issue was also promoted by staff with scientific training. The bill that moved last month was first drafted by a scientist working for Senator Frank Lautenberg (Democrat, New Jersey) under the Sea Grant Fellowship programme. The programme, run by the National Oceanic and Atmospheric Administration, brings graduate students and recent PhDs in marine and aquatic fields to Washington for one-year placements on Capitol Hill or federal agencies. Science fellows, in effect, increase the permeability of that congressional 'membrane' by offering more sites for binding.

But finding a willing audience and speaking clearly are not enough. Scientists spreading the word on acidification have been careful to maintain their credibility by being open about uncertainties and by drawing a line between science and policy. "The chemistry is highly certain," Caldeira says. "But the biological consequences are highly uncertain." His goal, he says, is to make clear what the risks are.

Even more importantly, COMPASS emphasizes that scientists should clearly distinguish between when they are describing science and when they are advocating what to do about the problem. Caldeira agrees: "I think that, as scientists, we have the ability and the right, if not the obligation, to speak as concerned and informed citizens. But it is useful to keep those roles separate. We have no particular priestly role where we have greater weight than anyone else" when it comes to policy-making.

Although the acidification story provides a model of how to get science on the congressional agenda, the tale is far from having a happy ending. As citizens, Caldeira and his colleagues want action to counter the problem. And that will require a debate engaging the entire Congress on matters far beyond science. Washington is edging ever closer to that, but it will require more than learning by osmosis. ■

David Goldston is a visiting lecturer at Harvard University's Center for the Environment. Reach him at partyofonecolumn@gmail.com.

*The conference history is documented in a paper by Gabrielle Dreyfus, a PhD student at Princeton University.

Battlefield insectica

Entomologists are briefing the military on how to protect troops from the scourge of the desert: sandflies. **Rex Dalton** reports.

The desert to the east of Palm Springs, California, is heaven for entomologists. Between date-palm fields irrigated with standing water and the inland Salton Sea buzz a bonanza of insects. Researcher Kenneth Linthicum makes his way through the low scrub, which his team has sprayed in places with the insecticide bifenthrin. He checks traps that spew carbon dioxide, a lure for mosquitoes.

Linthicum is on a quest to battle these pesky creatures, particularly in vulnerable populations such as US troops stationed in Iraq and other desert areas. He is one of a handful of researchers on the front lines of an entomological battle, working to reduce the incidence of insect-borne diseases among the military. In particular, they are targeting the tiny sandfly (genus *Phlebotomus*), which transmits the protozoan that causes the sometimes deadly disease leishmaniasis.

Leishmaniasis is a major public-health concern: 12 million people are infected worldwide, with an estimated 2 million new infections each year. Military researchers have taken the lead before in developing insect repellants that end up benefiting civilians; DEET, for instance, was developed as a consequence of World War II jungle battles. Now, with troops in Iraq and Afghanistan, scientists are turning their attention to the little-understood plague of sandflies. “It’s amazing we didn’t know more about this bug,” says Linthicum, who directs a US Department of Agriculture (USDA) research centre in Gainesville, Florida.

Irritating bites

When US troops invaded Iraq in 2003, soldiers woke regularly with a rash of highly irritating bites they called Baghdad boils. “Soldiers were getting eaten alive by the sandflies,” says Russell Coleman, a medical epidemiologist with the US Army who was deployed with the first troops at Tallil airbase in southern Iraq. “In certain units, 70% were taking dozens to hundreds of bites a night.”[1] More than 1,450 US troops have contracted leishmaniasis in the past four years, mainly in Iraq and Afghanistan.

In 2004, the military established the Deployed War-Fighter Protection Program, with a particular focus on sandflies. It receives US$5 million a year for projects overseen by the Armed Forces Pest Management Board in Washington, DC.

The programme includes scientists from all the services (Army, Navy, Air Force and Marines), along with academics and USDA researchers. They collaborate with military researchers from other countries, such as an Australian team who are studying ways to improve uniforms against insect bites[2]. “We are hoping to use the combined expertise to find new methods to protect personnel,” says Stanton Cope, a Navy entomologist based in Washington DC, who serves as a point of liaison between the researchers and the pest-management board.

A uniform approach

Leishmaniasis has two main forms: cutaneous, which creates skin lesions or attacks the mucous membranes, and internal or visceral, which can lead to liver and other organ problems and, in rare cases, death. Sandflies present a whole new challenge to military entomologists who are used to fighting mosquitoes. That’s because the insecticide usually used on military clothing, permethrin, often isn’t effective against the sandfly. “A mosquito may land on a uniform, bounce around, then fall dead from the insecticide,” says Ulrich Bernier, a chemist working at the USDA’s Gainesville centre. “A sandfly will be knocked out — then it gets up five minutes later.”

Uniforms are the first line of defence against bug bites. In the past, they were made of wool or cotton, with simple but durable weaves saturated in an insecticide such as permethrin. Today, uniforms have specially designed weaves of specific thickness in a mix of threads such as nylon, rayon and cotton to make them durable and protective. They are meant to protect against mosquitoes, ticks and fleas; but sandflies can still find their way through the cloth and bite — and that’s if the soldier has the uniform on in the first place, which is not common during summer nights in the desert. A further complication is that each of the four services has its own uniform, each with a different ability to guard against insects.

In the early 1990s, as the first Gulf War was starting, some services began issuing new uniforms pre-treated with insecticide. This

ARMED FORCES PEST MANAGEMENT BOARD

Bites (left) from the sandfly (right) expose US troops to the potentially deadly disease leishmaniasis.


J. RAEDLE/GETTY IMAGES

Science is guiding uniform design to prevent troops from insect bites while serving in desert areas.

barrier would be boosted periodically in the field by adding chemicals to the uniform. Then in 2003, a deployment of about 75 marines showed how vulnerable a soldier in uniform can be. Nearly 40% of the marines stationed in Monrovia, Liberia, during a civil disturbance contracted malaria. That's when marine commanders turned to the USDA centre to identify the uniforms' weakness.

Researchers there realized that the corps put their uniforms through a permanent press treatment to make them look ironed and the creases look sharp. The marines were the only service to do so. They had also switched to a twill fabric from a mixed weave. Those two factors let the mosquitoes through to bite, says Bernier.

New system

That was enough to get the corps to change their uniforms. "The marines raised the bar," he says. "They wanted to stop 100% of the biting." Working with USDA researchers and those at the Army Soldier Systems Center at Natick in Massachusetts, the marines launched new uniforms in March 2007. The garments now come pre-treated with a stronger insecticide so as to theoretically not need field treatments during the life of the clothing. Bernier is collecting uniforms from marines in Iraq to see how the new system is working.

One challenge, however, is to not saturate the uniform so heavily with insecticide that it starts to become toxic to the wearer[3]. Another complication is that the Army now treats its rayon/nylon uniforms with a fire retardant, as the earlier version would melt into wounds of those burned in bomb attacks. The question remains: how does that fire retardant interact or otherwise affect insect treatments on the same uniform?

> "Soldiers were getting eaten alive by the sandflies." — Russell Coleman

What's more, any changes to uniform designs are costly to implement. Marines go through about 385,000 uniforms a year, the Army 4 million. So beyond uniforms, researchers are looking at new and different methods for protecting soldiers from insects.

One approach is to better identify the chemicals that attract insects to humans, and use those to mask skin against the insects. Bernier's team, for instance, has identified at least three chemical blends that attract insects, based on chemicals that humans emit naturally[4].

Another approach is to develop entirely new types of insecticides. Chemist Alan Katritzky of the University of Florida in Gainesville reported recently on an artificial neural network that they used to screen mosquito repellants[5]. His computer modelling system used past data on mosquito repellency to identify 34 new candidate repellants. When tested in the lab, some of the candidates repelled mosquitoes up to three times as long as DEET.

At Louisiana State University in Baton Rouge, researchers are targeting not just the sandflies but also the rodents in whose burrows the flies live. Scientists are studying a feed mixed with a pesticide (novaluron) for rodents. The pesticide is passed onto immature sandflies when they eat the rodents' faeces[6]. This could help solve another problem: troops bulldozed entire areas where tents were being erected to get rid of sandflies, but more would later pop up from the rodent burrows underneath.

Precision spraying

Meanwhile, USDA researchers in College Station, Texas, are developing a system in which a small, unmanned helicopter is programmed via a computer to spray a set area from 22:00 to 4:00, key hours for sandfly activity. "We mounted a sprayer and global positioning system (GPS) on an expensive toy helicopter," says USDA agricultural engineer Clint Hoffmann. "You set the helicopter's flight pattern with a computer mouse." Field testing is coming soon, with the first overseas study planned for Egypt.

Even the size of droplets sprayed is receiving scrutiny[7]. If you're trying to hit a flying insect, you want a 25-micrometre droplet. If you want to cover plants or tents, a 100-micrometre droplet is best. "We have a fairly sophisticated drift model that can track individual droplets — even through aircraft vortices or wakes," Hoffmann says. These systems have shown that the standard methods the military used to trap and kill insects, such as spraying in the early evening when sandflies aren't around, are virtually worthless.

It may take a while for the research to work its way into military consciousness, and eventually on to benefit civilians. But with the way the war is going in Iraq, there could be plenty of time left to learn the lessons. ■

Rex Dalton writes for *Nature* from San Diego.

1. Coleman, R. E. *et al. J. Med. Entomol.* **43,** 647–662 (2006).
2. Frances, S. P. & Cooper, R. D. *ADF Health* **8,** 50–56 (2007).
3. Macedo, P. A. *et al. J. Toxicol. Environ. Health A* **70,** 1758–1771 (2007).
4. Bernier, U. R. *et al. J. Med. Entomol.* **40,** 653–656 (2003).
5. Katritzky, A. R. *et al. Proc. Natl Acad. Sci. USA* **105,** 7359–7364 (2008).
6. Mascari, T. M. *et al J. Med. Entomol.* **44,** 714–717 (2007).
7. Hoffmann, W. C. *et al. J. Am. Mosq. Control Assoc.* **23,** 315–320 (2007).

DETAILS OF BEING HUMAN

A difference in one molecule led physician Ajit Varki to question what sets humans apart from other apes. **Bruce Lieberman** meets a man who sees a big picture in the finer points.

B. LIEBERMAN

The human body does not welcome an injection of horse serum. Ajit Varki discovered this when, as a young San Diego doctor in 1984, he administered some to a woman with bone-marrow failure. The serum was a standard treatment intended to stop the woman's T cells from destroying her bone marrow. But it was also known to prompt a reaction called 'serum sickness' and, sure enough, the patient broke out in hives a week after treatment — the result, Varki assumed, of her immune system's assault on proteins from another species.

Soon after observing his patient's reaction, Varki learned that proteins weren't the only thing to blame. So were sialic acids, sugars that carpet the surface of mammalian cells. Some studies had suggested that the human immune system reacted against one sialic acid called *N*-glycolyl neuraminic acid (Neu5Gc) in the horse serum. "How can that be?" Varki remembers thinking. "How can you have a reaction against sialic acid? It's everywhere. All mammals have sialic acid." Varki wondered whether humans might in fact be the only mammal that lacked Neu5Gc.

A physician and biochemist by training, Varki had already embarked on a career in the relatively new field of glycobiology, the study of the sugar chains that decorate many proteins and lipids inside and outside the cell. But it was another 14 years before he got the chance to answer his original question. In 1998, he and his colleagues used high-performance liquid chromatography to analyse blood samples from chimps, bonobos, gorillas, orangutans and humans. They found that humans are indeed the only primates missing Neu5Gc[1] and that human cells are instead rich in another sialic acid, *N*-acetyl neuraminic acid (Neu5Ac).

A career in evolution

These findings started Varki off on a road that led to his becoming not only a leading glycobiologist but a respected 'honorary' palaeoanthropologist. He is one of the co-founders and directors of the multidisciplinary Center for Academic Research and Training in Anthropogeny (CARTA) — a research collaboration between the University of California, San Diego, and the Salk Institute in nearby La Jolla. The centre was launched in March this year with a US$3-million grant from the G. Harold & Leila Y. Mathers Foundation, based in New York state.

> **"What we're dealing with here is a gene loss that has an effect throughout the whole body."**
> — Morris Goodman

The 'Anthropogeny' in the centre's title resurrects a term for the study of both the evolution and the individual development of human beings that would have been familiar to earlier generations of anthropologists. To Varki, the word encapsulates some of the biggest questions in the study of human origins, such as how, why and when the human brain evolved its present functions. One of his latest research projects is a collaboration with Spanish palaeontologist Juan Luis Arsuaga, of the Complutense University of Madrid, for the biochemical analysis of 900,000-year-old *Homo antecessor* fossils from Atapuerca in northern Spain, some of the oldest hominid bones yet found in Europe. What Varki is looking for is evidence that Neu5Gc was lost very early in human evolution. He believes that the fact that humans, and only humans, have lost Neu5Gc could be implicated in the emergence of hominid species.

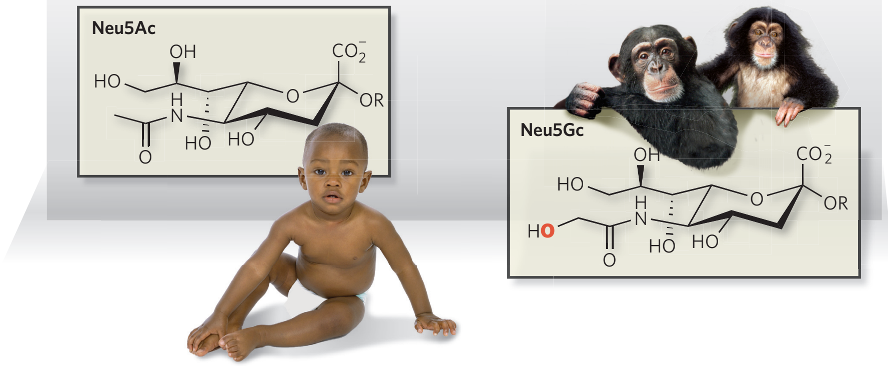
The journey from glycobiologist to director of a multidisciplinary human origins centre has been fuelled by Varki's insatiable desire for knowledge. "The guy is just an encyclopaedia," says glycobiologist Mark Lehrman at the University of Texas Southwestern Medical Center in Dallas. "Even though he wasn't trained in anthropology, he's been able to educate himself in this area and become an authority. It's a remarkable gift to be able to do that and do it well."

Varki initially trained as a general medical doctor at the Christian Medical College in Vellore, India. To pursue a dual medical and research career, he went to the United States, eventually taking up a fellowship under Stuart Kornfeld at Washington University in St Louis, Missouri, in the late 1970s.

Kornfeld was beginning his work on sugar chains, including sialic acids, and Varki was intrigued by the opportunity to contribute to a largely unexplored area of biology. In 1982, he set up his own glycobiology lab at the University of California, San Diego, where he still works today.

SMALL CHANGE, BIG DIFFERENCE

A mutation during human evolution means that humans accumulate the sialic acid Neu5Ac whereas other primates also make Neu5Gc.



On a molecular level, the difference between Neu5Gc and Neu5Ac is tiny — a single added oxygen atom perched on one arm distinguishes one from the other (see graphic). But on a biological level, the difference could be enormous. "We thought if monkeys and all of our closest relatives have Neu5Gc and humans don't, then there must be a molecular basis for that," Varki says. He subsequently found it in an enzyme that converts Neu5Ac to Neu5Gc, but which is disabled by mutation in humans[2].

Selection pressure

Varki's discovery pointed to a definitive difference that set chimps and humans biochemically apart, says Morris Goodman, an evolutionary biologist at Wayne State University in Detroit, Michigan. It was one of the first such differences to be found, and because sialic acids serve many biological roles, primarily as cell-recognition and cell-adhesion molecules, it might explain some of the unique aspects of human biology. "What we're dealing with here is a gene loss that has an effect throughout the whole body," says Goodman.

At the time, Varki realized he knew little about human evolution except what he'd learned as an undergraduate or read in *National Geographic*. So he set out to educate himself. He took a short sabbatical at the Yerkes National Primate Research Center in Atlanta, Georgia. Reviewing the animals' medical records with a veterinarian, he learned that the centre had never seen a case of rheumatoid arthritis or bronchial asthma — common conditions in humans. Chimpanzees don't get sick from the human malaria parasite, *Plasmodium falciparum*. Conversely, humans can't be infected with *P. reichenowi*, the malaria parasite that plagues chimpanzees.

In subsequent work, Varki and his team showed that the different susceptibilities were due to the differences in sialic acids. *P. reichenowi* prefers to grab hold of Neu5Gc on chimp red blood cells, whereas *P. falciparum* favours Neu5Ac[3]. The researchers hypothesized that the selection pressure to evade *P. reichenowi* may have led humans to lose Neu5Gc and acquire resistance to this parasite — and that this loss may have helped to fuel the emergence of *P. falciparum*, which could gain entry by latching onto Neu5Ac instead. Other discoveries in Varki's lab — including ten other human-specific genetic changes affecting sialic acid function — may help to explain uniquely human vulnerabilities to conditions such as Alzheimer's disease and multiple sclerosis.

J. TRUEBA/MSF/SPL

Tests of *Homo antecessor* fossils could help confirm the loss of Neu5Gc early in human evolution.

Varki's interest in human evolution soon extended far beyond chimps and their sugars. "I found he was talking with several people on campus," says neuroscientist Fred Gage at the Salk Institute, a long-time collaborator and friend. "I told him that it wasn't fair that

he would have these one-on-one conversations and not share what was being talked about," he jokes.

Reimagining anthropogeny

Gage encouraged Varki to organize a series of informal seminars on human origins at the university. Between 1998 and 2007, the Project for Explaining the Origin of Humans drew in anthropologists, primate biologists, geneticists, immunologists, neuroscientists, linguists and many others. They discussed topics ranging from the evolution of language to the differences between humans, Neanderthals and *Homo erectus*, the first hominid to leave Africa. Goodman says the interdisciplinary nature of the series made it extremely important to the field. "You really had the chance to explore an issue as it relates to the evolutionary origins of our species," he says.

Varki's motivations were partly selfish: "One of my goals, my secret agenda, was to educate myself," he admits. "At the last meeting I asked the people who attended if I could have a bachelor's degree in anthropogeny." Varki estimates that he has listened to more than 300 talks on various aspects of this discipline. "The idea is the linguist needs to talk to the molecular biologist who needs to talk to the neuroscientist who needs to talk to the psychologist and philosopher about these issues," he says. "Most areas of human knowledge are somewhere relevant."

CARTA is a successor to the human origins series. Directed by Varki, Gage, Margaret Schoeninger, a professor of anthropology at the University of California, San Diego, and Pascal Gagneux, a primate biologist and Varki's close collaborator, the centre already has some 40


P. TWEEDIE/CORBIS

Differences in sialic acids between chimps and humans alter susceptibilities to some diseases.

San Diego-based members and more than 100 in the rest of the United States and elsewhere in the world.

CARTA aims to foster connections between these researchers worldwide, facilitate access to resources for great-ape research, develop a peer-reviewed journal and offer courses on human origins. The project is in some ways comparable to the Leipzig School of Human Origins in Germany, an interdisciplinary PhD programme run jointly by the Max Planck Institute for Evolutionary Anthropology in Leipzig and Leipzig University since 2005. Varki says that CARTA will be more of a virtual organization and that "the effort should transcend disciplines", pointing as an example to his own work on sialic acids, which has required collaboration between biochemists, palaeontologists and physicians.

Acid test

Back in the lab, Varki and Gagneux will in the next few months embark on the preliminary analysis of animal fossils from Atapuerca, to see if they can detect preserved sialic acids using high-performance liquid chromatography and mass spectrometry. If so, sialic acids are likely to be preserved in hominid fossils from the same strata and the researchers will test those next.

"Palaeontologists are usually seen as people interested in something that is finished and belongs to the past," Arsuaga says, "and usually the idea is missed that we are looking for an explanation of living humans." He says he was persuaded to let tests be done on the precious *H. antecessor* fossils because "the damage is not big" from current techniques that drill small amounts of powder from inside the bone.

Varki and Gagneux hope that these fossils may help to answer some grand hypotheses about Neu5Gc and its role in human evolution. They estimate that the mutation that caused the loss of Neu5Gc first appeared among human ancestors 2 million to 3 million years ago, which coincides with the emergence of *H. erectus*, and they believe that pathogens such as malaria may have initiated this change. They wonder whether the change in this ubiquitous sugar could have had other broad-ranging biological effects that helped create reproductive isolation between those with Neu5Gc and those without, and whether these effects could have contributed to the emergence of *H. erectus*, followed by *H. antecessor*. "Losing Neu5Gc may have been great for survival, but it may have forced you to forgo reproduction with a whole group of your former buddies who didn't undergo this change," Gagneux says. If they can show that Arsuaga's *H. antecessor* fossils also lack Neu5Gc, this will be yet more evidence in support of their hypothesis.

"Understanding where we came from is very important to understanding where we're going."
— Ajit Varki

If ancient humans can't answer the speciation hypothesis, then perhaps mice will help. Varki and Gagneux have genetically engineered mice that lack the Neu5Gc sialic acid that humans are missing and Varki says that they display subtle human-like features[4]. Compared with wild-type mice, they have poor hearing, somewhat reminiscent of human age-related hearing loss, and slower wound healing, as do humans compared with non-human primates. Further studies should reveal whether these mice are able to reproduce with wild-type animals that still have Neu5Gc.

Varki's recent work has brought him back to the immune reaction he observed nearly 25 years ago. Even though humans don't make Neu5Gc, it is eaten in animal products that contain it, such as meat and milk. Varki and Gagneux wonder whether — among meat-eaters at least — Neu5Gc elicits an immune reaction that might contribute to a whole spectrum of human-specific diseases that are associated with chronic inflammation, including heart disease and cancer. Such diseases would not have been such a problem when humans had shorter life spans.

Food for thought

To test the idea, Gagneux took a trip to a local Whole Foods Market, loaded up a shopping cart with meat and dairy products and took them back to the lab for analysis. The researchers found the highest levels of Neu5Gc in lamb, pork and beef. "We swallowed big bowls of that and we collected every possible sample we could from ourselves in the following few weeks to see whether it shows up in our own glycoproteins," Gagneux says, "and the answer is yes, it does." The team has also found that many people carry antibodies targeted against the sugar[5].

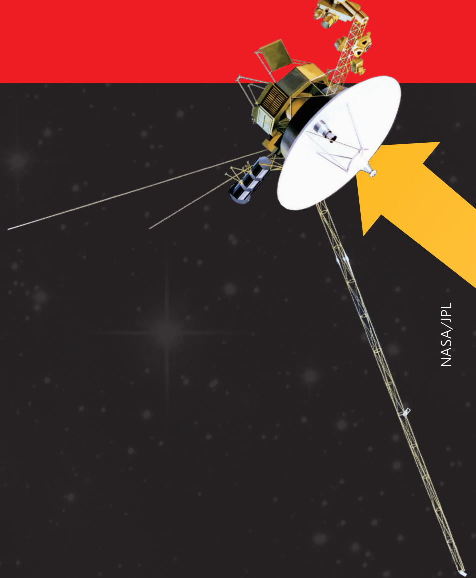
If their hypothesis holds up, it will illustrate how selection pressures change: where once selection favoured the loss of Neu5Gc to protect hominids from pathogens, now its absence could be making humans susceptible to other diseases. "Once you've lost it, you have to make do with what you have," Varki says.

For Varki, who began his professional life observing patients, these studies have brought him full circle. The molecules that made humans human may be the same ones that make us uniquely vulnerable to our most threatening diseases. "In some cases, they would be what I call the scars of our evolution," Varki says. "My experience has opened my mind to the fact that understanding human evolution, where we came from, is very important to understanding who we are and where we're going." ■

Bruce Lieberman is a freelance science writer based in San Diego.

1. Muchmore, E. A. *et al. Am. J. Phys. Anthropol.* **107,** 187–198 (1998).
2. Chou, H.-H. *et al. Proc. Natl Acad. Sci. USA* **95,** 11751–11756 (1998).
3. Martin, M. J. *et al. Proc. Natl Acad. Sci. USA* **102,** 12819–12824 (2005).
4. Hedlund, M. *et al. Mol. Cell. Biol.* **27,** 4340–4346 (2007).
5. Tangvoranuntakul, P. *et al. Proc. Natl Acad. Sci. USA* **100,** 12045–12050 (2003).

NATURE|Vol 454|3 July 2008



NASA/JPL

VOYAGER 1

LAUNCH: SEPTEMBER 1977
JUPITER ENCOUNTER: MARCH 1979
SATURN ENCOUNTER: NOVEMBER 1980
OVERTAKES PIONEER 11 TO BECOME FARTHEST SPACECRAFT FROM EARTH: FEBRUARY 1998
ENTERS HELIOSHEATH: DECEMBER 2004
SCIENCE INSTRUMENT SHUTDOWN BEGINS: 2020

What a long, strange trip it's been

Launched in 1977, NASA's Voyager missions transformed humanity's view of the Solar System. Now in their fourth decade, they are sending back information about the borderlands of interstellar space (see pages 63–83). Here, three veterans recall details and moments that meant something special along the way.

The little motor that could

JOHNS HOPKINS APPLIED PHYSICS LAB.

We needed an instrument that scanned the sky in order to determine the direction that particles flying past the spacecraft were travelling in. But how do you get an instrument to scan 360° of the sky when the spacecraft on which it's mounted doesn't rotate, and must always point to Earth?

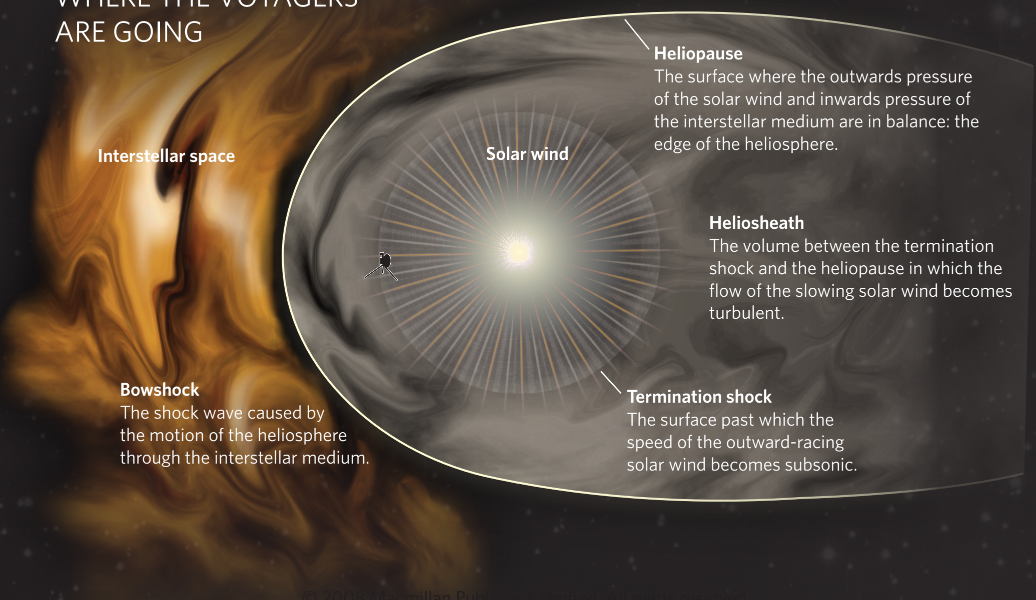
Our answer was to mount the instrument on a small platform and have an electrical stepping motor move it 45° every minute or two — sometimes a lot faster. The engineers, being a conservative bunch, thought that after a few thousand steps the motor would seize up. Remember, it was 1971! A design that involved moving parts on a spacecraft was not considered a good idea — the initial response was 'Are you crazy?'. I insisted, and we tested the motor for about 500,000 steps, which was twice what we thought we'd need for the trip to Saturn.

Over 30 years on and after well over 5 million steps the little motor is still stepping dutifully once every 192 seconds. That's 10 million steps if we count Voyagers 1 and 2. Not only did we get wonderful directional data as we went through the magnetospheres of four planets, but the direction of particle flow signalled the approach to the termination shock a couple of years before we actually got there!

It was the riskiest decision I ever made, but it paid off big time, scientifically. For a scientist, the Voyager mission is the stuff that dreams are made of.

Stamatios Krimigis has been principal investigator of the low-energy charged particle experiments since 1971.

WHERE THE VOYAGERS ARE GOING



BACKGROUND: NASA, ESA, HUBBLE HERITAGE TEAM (STSCI/AURA)-ESA/HUBBLE COLLABORATION

For more information on the Voyager missions, see the News and Views article on page 38; the sequence of six Letters starting on page 63; the video at www.nature.com/nature/videoarchive/voyager; and the podcast at http://www.nature.com/nature/podcast.

The cup that cheers

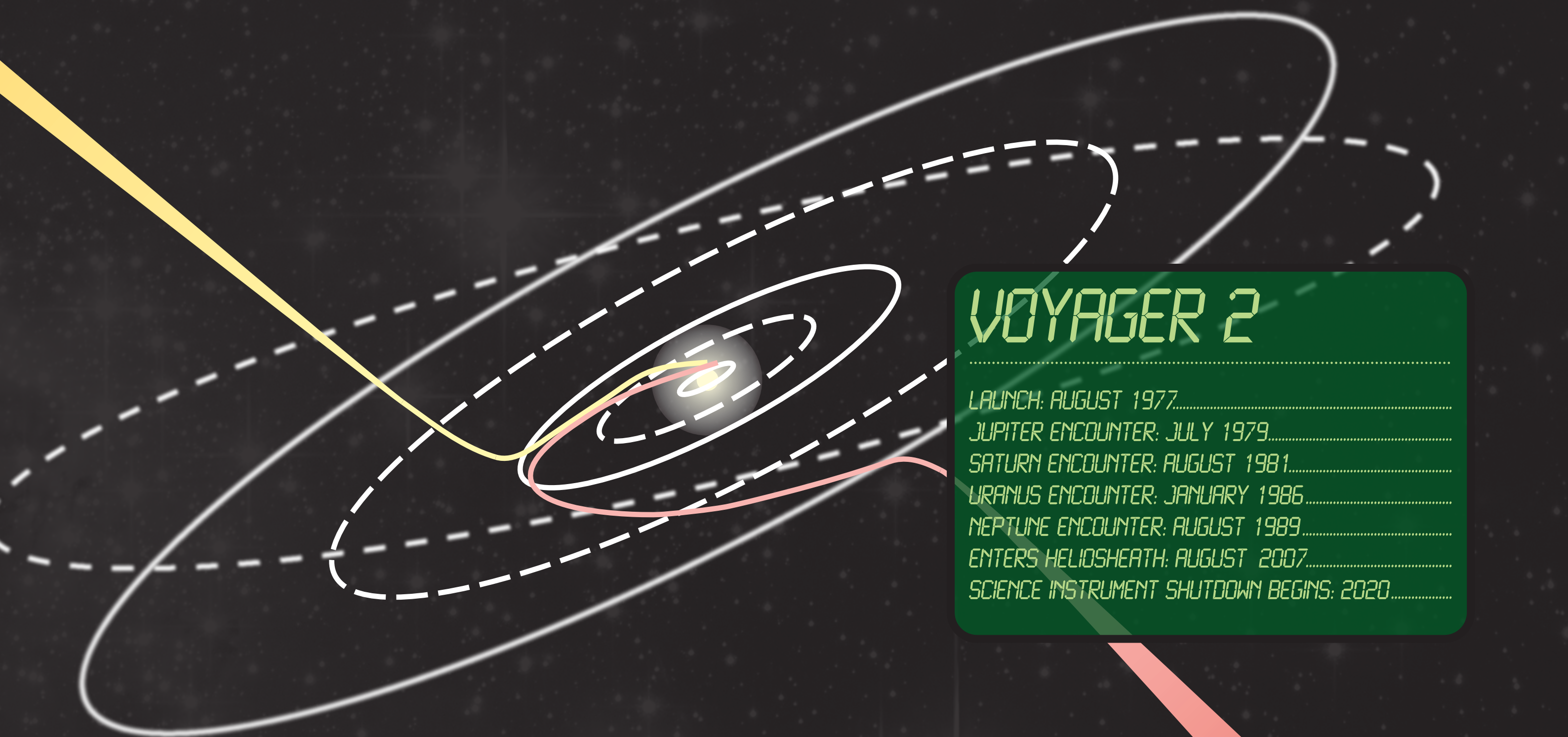

D. STAHLMAN

The instrument concept for the Plasma spectrometer — the 'Faraday cup' design — dates from the beginning of the space age. The late Herb Bridge, my mentor at the Massachusetts Institute of Technology, used something very similar to make the first measurements of the energy spectrum of the shocked solar wind in the earth's magnetosheath in 1961, when he was principal investigator on the plasma instrument on Explorer 10.

And there's an excellent chance that, before I retire, the Faraday cup on Voyager will directly measure the plasma in the interstellar medium. I was in high school when Explorer 10 went up, and in my lifetime we have gone from being just above the atmosphere to reaching for the interstellar medium. Pretty nifty.

Everyone knows about the golden records on the Voyagers that Carl Sagan worked on. Our cup collector has a message too. The names of all the people at MIT who worked on the plasma probe are written on the surface underneath the motto "Live Free or Die". Our chief engineer on the project was Bob Butler, a New Hampshire resident, and that's the New Hampshire state motto.

John Belcher was principal investigator of the plasma spectrometer from 1986 to 1990.



The mission of a lifetime

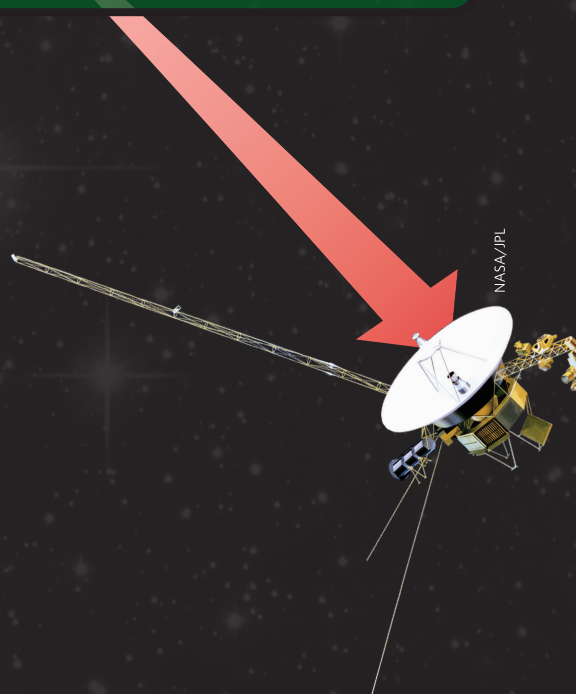

D. PERRY, UNIV. DELAWARE

The most exciting moments for me were the six close planetary encounters, especially Voyager 2's discovery of previously unknown and unpredicted but significant intrinsic global magnetic fields at Uranus and Neptune, with their entrapped radiation belts. It was surely as exciting as discovering the magnetic field of Mercury in 1974 and 1975 with Mariner 10, or discovering the complexly magnetized crust of Mars in 1999 with the orbiting Mars Global Surveyor.

For Voyager 2's Uranus fly-by in 1986 the magnetometer team invited James Van Allen, discoverer of Earth's radiation belts in 1958, to join us at the Jet Propulsion Laboratory for the encounter activities. He was as delighted to participate in the real-time learning episode as if he were a fellow principal investigator.

Over the years the magnetometer team has been downsized by 50%, mainly by recent retirements. But along with long-time NASA Goddard Space Flight Center colleagues Mario Acuña and Leonard Burlaga we soldier on as the data continue to flow in. It continues to be the mission of a lifetime.

Norman Ness has been principal investigator of the magnetometers since 1971.


NASA/JPL

CORRESPONDENCE

Amateurs as an outreach of HAARP's lunar-echo study

SIR — Your News Feature 'Heating up the heavens' (*Nature* **452,** 930–932; 2008) discusses experiments using the High Frequency Active Auroral Research Program (HAARP) facility. I would like to clarify the goal of the lunar-echo experiments.

The high power and low-frequency range of HAARP provide radio waves that can penetrate the lunar surface because of the low electrical conductivity of the lunar regolith, allowing investigation of the subsurface. Parallel research efforts currently use radar sounders on satellites orbiting the Moon. In its radar mode, HAARP is the most powerful of the few Earth-based facilities that can participate in such investigations.

Our lunar-echo experiments began in 2001, when HAARP was not yet at full power capability, but in the latest experiment we succeeded in receiving a lunar echo at the lowest frequency obtained so far by an Earth-based radar (4.8 megahertz). Our initial results will be presented at the General Assembly of the International Union of Radio Science in August 2008.

The fact that radio amateurs can record the HAARP transmissions and lunar echoes is a result of free-space propagation of the radio waves. In the lunar-echo experiment of January 2008 that you mention, the schedule of the HAARP transmissions was provided so that they could listen in during the research activities. That opportunity was enthusiastically and successfully taken up by radio amateurs in many countries and has produced interesting data in its own right, related to world-wide propagation effects on the two-way (Earth–Moon–Earth) radio signals — for example, on ionospheric refraction, scattering and scintillation.

Paul Rodriguez ITT Corporation Consultant, Code 5550, Naval Research Laboratory, Washington DC 20375, USA

Reality lags behind rhetoric in building interdisciplinary work

SIR — As a PhD student in archaeology and genetics, I am all too aware of the difficulties in crossing a gaping discipline divide, as well as of their effect on academic career prospects, as discussed in the *Naturejobs* article 'Assembly work' (*Nature* **453,** 422–423; 2008).

For my master's degree in biomolecular archaeology, I needed a foot in two UK universities: one in the University of Manchester's biology department and the other in the University of Sheffield's archaeology department. My former lecturers later became part of the Manchester Interdisciplinary Biocentre (MIB) and the MSc course shifted to Sheffield, where eventually the programme ceased.

This closure was a disappointment for the nascent field of bioarchaeology, set to thrive only on a foundation of solid postgraduate training. Although the MIB and other new centres for interdisciplinary research are enthusiastically welcomed, they are few and far between and so able to offer only limited postdoctoral prospects.

Opening such centres and creating training programmes is not enough — it is also necessary to make interdisciplinary fields attractive to graduates and for senior academics to appreciate their significance. This would improve project turnover, bringing more funding to collaborative projects that would sustain interdisciplinary centres and allow academics from each discipline to gauge publications on an equal footing.

Perhaps then my search for a postdoctoral position in bioarchaeology would be easier. Although interdisciplinary projects are viewed as hot topics, in reality they lag behind as they await official establishment and recognition.

Danae Rebecca Dodge Graduate School of Archaeology, West Court, 2 Mappin Street, University of Sheffield, Sheffield S1 4DT, UK

Reduce confusion by using 'design' more intelligently

SIR — Few scientists would dispute that evolution provides a far more satisfactory explanation for the workings of living organisms than does 'intelligent design'. But a much more subtle 'design' movement abounds that can distort how they approach their research.

According to the *Oxford English Dictionary*, the noun 'design' means "the thing aimed at; the end in view; the final purpose". Biologists frequently use 'design' to describe the organization of components in a system — for instance, in a network of signalling molecules, a cell's cytoskeleton, or a protein's domain structure. Readers need look no further than articles in scientific publications and their podcasts for other examples.

It would be easy to dismiss criticism of this as mere pedantry, if it weren't for puzzled researchers voicing concern at the seemingly illogical or counterintuitive 'design' of a biological structure or process. They fail to note that its construction is a consequence of the incremental, biased random walk of evolution, and this can affect their choice of approach in tackling the problem.

Systems that emerge by selection differ fundamentally from those conceived by design. Failing to acknowledge this in our choice of words is lazy, clutters our thinking and does a disservice to those entering biology from disciplines (scientific and non-scientific) in which evolution by selection is not a central theme and the word 'design' carries inherent baggage.

Richard Sever Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA

Picture not quite worth 1,000 words in this case

SIR — In your News story 'Top billing for platypus at end of evolution tree' (*Nature* **453,** 138–139; 2008), the graphic depicting genome status presents a shocking new phylogeny of the Vertebrates — with Archosaurs (birds and crocodilians) and Mammals forming a monophyletic group.

The bad news is that this dramatic new 'proposal' is completely adrift from the research Article by Wesley C. Warren and colleagues that the figurative tree is intended to illustrate (*Nature* **453,** 175–183; 2008), and it continues a persistent tendency in popular literature to portray all evolution as leading towards humans.

The good news is that figures 1 and 4 in the original Article both correctly show the properly supported phylogeny that holds the Archosaurs as sister group to the Turtle + Diapsid clade. Even better news is that a news story that appeared in *Science* the following day (*Science* **320,** 730; 2008) about the platypus research also showed a cladogram containing essentially the same error. One wonders whether the similarity in errors between the two news stories is an example of homology or homoplasy.

Peter K. Ducey Department of Biological Sciences, State University of New York at Cortland, Cortland, New York 13045, USA

BOOKS & ARTS

Perils of perversity

Research is riddled with strong characters; **Walter Gratzer** applauds a spirited attempt to get their measure.

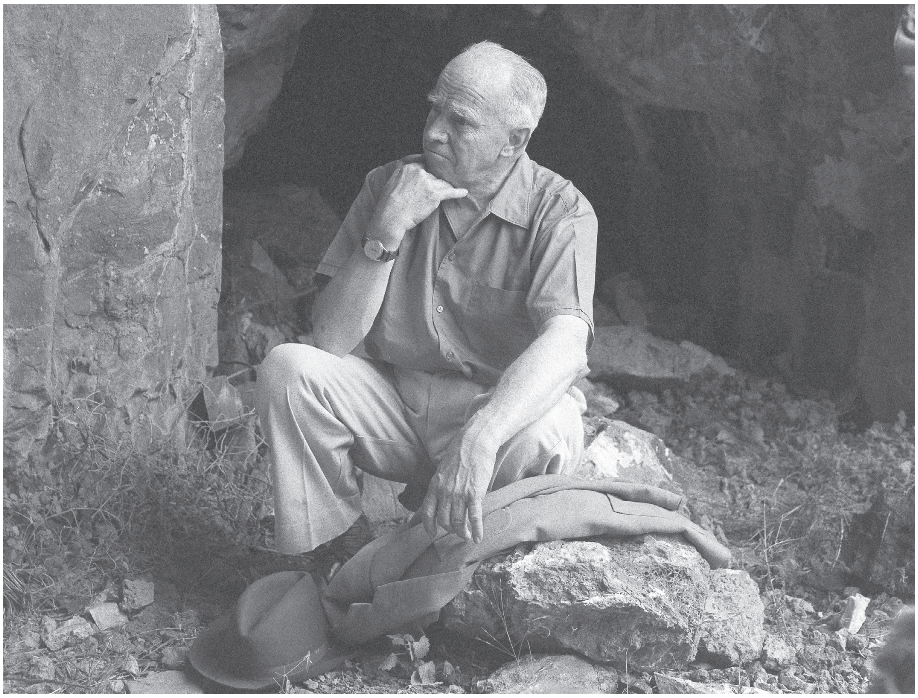
Rebels, Mavericks, and Heretics in Biology
Edited by Oren Solomon Harman and Michael R. Dietrich
Yale University Press: 2008. 352 pp. \$40.00, £25

The proposition put in the mouth of Lemuel Gulliver by Jonathan Swift that "there is nothing so extravagant or irrational which philosophers have not maintained for truth", still holds. The zealots, monomaniacs and obsessives at whose approach we scatter are in general just that, and nothing more. For it is hard for the free-ranging intellect to walk the fine line between scepticism, the stuff of science, and credulity, which would drown us in a deluge of homoeopathy, parapsychology, perpetual-motion machines and 'intelligent design'.

This collection of essays, each about a scientist held by the editors Oren Harman and Michael Dietrich to qualify for the accolade of rebel, maverick or heretic, aims at a definition of the species. Even if little emerges in the way of broad insights into these traits, most of the narratives make compelling reading, for the contributors manage to link the science to the personalities of the protagonists and to the social and professional milieu in which they functioned.

Of the nineteen subjects, eight were students of evolution, and only three biochemists or molecular biologists, so the biological terrain is a trifle skewed. History has been kind to most of them, for their heterodox views have largely prevailed. I would have liked to hear a little about some, prominent or scandalous in their day, who have faded into oblivion. Paul Kammerer, for one, whose purported demonstrations of Lamarckian inheritance brought him adulation, before accusations of fraud drove him to suicide. And Kammerer's apologist, the egregious E. W. MacBride who presided over a distinguished biology department in London and whose noisy affirmations of Lamarckian evolution and spontaneous generation continued into the middle of the twentieth century.

The most felicitous chapters are by writers who worked with or knew their man or, in two cases, woman. Most stirring is the account of the life and labours of the Australian palaeontologist, Raymond Dart, by his friend, protégé and successor, Phillip Tobias. Dart's description of the Taung skull, with anatomical attributes of humans and apes, and his conclusion that it represented the earliest hominid, was greeted by his confrères with outrage, or at best disdain. A correspondent advised Dart that he would "roast in the quenchless fires of hell" for his pains; another that he was "a priest of Baal"; while a third predicted that Dart's destiny was to be "placed in an institution for the feeble-minded". Dart was made to feel the hauteur of the academic mandarins Grafton Elliott Smith, his former patron in Oxford, Arthur Keith and Arthur Smith Woodward. It is a pleasing irony that these pillars of the English establishment had not hesitated a few years earlier to authenticate the fraudulent Piltdown skull as the earliest hominid.


AP PHOTO

Palaeontologist Raymond Dart's description of the Taung skull outraged his peers.

Tobias suggests reasons behind the rancour. The Taung child had surfaced on the wrong continent, for according to the prevailing dogma, Asia was the seedbed of human evolution. This belief was allied to a conviction that nothing good could ever emerge from Africa. And there were technical questions, involving the favoured theories about teeth and other anatomical features, and dating. The whole field, moreover, was obfuscated by the Piltdown hoax, with its gratifying implication that early men with a large brains had sprung, like most good things, from the south of England.

Dart was pugnacious and opinionated, as were many of the other figures who stalk the pages of *Rebels, Mavericks, and Heretics in Biology*. Among them was C. D. (Cyril) Darlington, who maintained that his conception had been an accident, one that his father never risked repeating. He escaped from a claustrophobic family environment into an unpromising position as a volunteer assistant at the John Innes Horticultural Institute. Abrasive and unwilling to embrace the opinions of others, he managed nevertheless to get a toehold on the academic ladder, and soon discovered how chromosomes behaved during meiosis. Darlington's vision was viewed as a threat to the integrity of cell biology and an affront to its practitioners. A transatlantic researcher offered to cross the water and knock Darlington's head off, and at a lecture in the United States he was shouted down by an angry audience. Overtly at least, scientific discourse is conducted more decorously today.

J. B. S. Haldane, who flits in and out of these stories like a will-o'-the-wisp, saw the merits of Darlington's work and probably helped to advance his rock-strewn career. Genetics was

an exceptionally disputatious field. A schism had existed between a rigorous mathematical and the largely qualitative traditional approach since Francis Galton's time. Galton's successor at University College London, Karl Pearson, recalled indeed that, in 1900, when he submitted a paper to its *Proceedings*, "a resolution of the Council of the Royal Society was conveyed to me, requesting that in future papers, mathematics should be kept apart from biological applications". Pearson was sufficiently incensed to seek Galton's advice as to whether he should resign his fellowship. Darlington, by contrast, evidently relished the contumely.

Some of the biologists in this book scarcely qualify as rebels or heretics. They were first-rate scientists, sure of their data and of the obtuseness of their critics. Notable were Motoo Kimura, who established the principle of neutral evolution, Oswald 'Fess' Avery, who identified DNA as the genetic material, and Howard Temin, who discovered reverse transcriptase. Fine chapters — by James Crow, Kimura's PhD supervisor, by Ute Deichmann and by Daniel Kevles, who clearly knew and admired Temin as man and scientist — do them justice.

Barbara McClintock, on Nathaniel Comfort's evidence, was well liked too, despite portraying herself as swimming ever against the tide. She claimed to have been ostracized and thought "a little mad", when she first presented her observations on linkage groups. In truth, she was respected from the outset. Richard Lewontin in the book's epilogue relates an overheard conversation between two colleagues leaving a lecture: "That is the craziest thing I ever heard", one remarks, "but if Barbara says it, there must be something to it". The historical image diverges, as so often, from the reality.

Can one draw any moral from the sum of these excellent narratives? That the bloody-minded shall inherit the earth? Or perhaps merely the enduring principle of human endeavour, encapsulated by Charles Kettering of General Motors: "First they tell you you're wrong and they can prove it; then they tell you you're right but it isn't important; then they tell you it's important but they knew it all along". ■

Walter Gratzer is in the Randall Division of Cell and Molecular Biophysics, King's College, New Hunt's House, Guy's Campus, London SE1 1UL.

Bonding as key to hominid origins

Primatology meets socio-cultural analysis in a controversial account of human evolution.

P. BOWATER/ALAMY

A father in Papua New Guinea displays his marriageable daughters.

Primeval Kinship: How Pair-Bonding Gave Birth To Human Society
by Bernard Chapais
Harvard University Press: 2008. 368 pp. $39.95, £25.95, €30.00

Over the centuries answers to the question of what makes us uniquely human have varied. The Greek philosophers emphasized reason, Enlightenment scholars pointed to our political nature, and nineteenth-century historical evolutionists prioritized our complex social organization. After Charles Darwin and Sigmund Freud, answers became increasingly psychological, and in recent decades attention has turned to the genome and cognition. For the most part, modern social scientists eschew grand debates over human origins, fearful of becoming tarred with the brush of nineteenth-century social Darwinism.

Primeval Kinship returns to the big questions about ancient society that mesmerized the prominent social theorists of the Victorian era such as Edward Tylor, Lewis Henry Morgan and Friedrich Engels. The author, primatologist Bernard Chapais, offers a powerful and controversial new account of hominid origins. The root of humanity, he argues, lies in pair bonding (the strong affinity that can develop in a breeding couple), the brother–sister tie, and the transfer of females between groups. Imagine a violent chimp-like encounter between two groups of early humans. An attacking dad in one group recognizes his daughter in the other and thinks twice about killing her baby. Meanwhile the daughter's bonded mate recognizes his sister in the attacking group, and refrains from a counterattack. We might even see some reconciliatory grooming. Both males have a vested interest in the same mother–offspring pairs, and hence, to cut Chapais' intricately woven argument short, females act as peacemakers between groups.

A unique feature of human society is the largely peaceful relations between tribes. Chapais reiterates and builds upon the central claim in anthropologist Claude Lévi-Strauss' 1949 classic treatise *Les Structures Élémentaires de la Parenté* about what makes us human. Namely, that men exchange daughters and sisters between kin groups to form alliances. The idea goes back to Tylor's aphorism "marry-out or be killed-out". Placing the story in its phylogenetic context, Chapais radically reorders some of Lévi-Strauss' logic and develops a more fully fledged thesis. For instance, in contrast to Lévi-Strauss' claim, hominids did not invent incest taboos to achieve marital exchange, but inherited them from much deeper animal roots; outmigration of one sex (exogamy) was already in place when the last common ancestor of humans and chimpanzees stalked the forests.

Chapais' key contribution is to ask how hominids could recognize relatives who had transferred to other groups. He presents evidence from macaques and other primates that mother–child recognition, and cognizance of other mother–child dyads, can exist without the long-term establishment of sexual bonds between mates. Father–offspring recognition, by contrast, requires relatively durable pair bonds. Individuals who can identify a father–child relationship in others can more confidently identify their own siblings. Then, his logic goes, if one sex migrates to other groups, these long-distance sibling links will form the kernel of peaceable intergroup relations. Chapais stands apart from his eminent anthropological forebears, such as Leslie

White and Robin Fox, all of whom posited that language was the key to exogamy.

Primeval Kinship makes several claims that will upset biological anthropologists. Chapais' model requires that ancestral hominids lived in male-centred kinship groups. Essentially he proposes a nuanced homology between several systems in which males stay in their natal groups and females transfer — in chimps or bonobos and in hunter–gatherers. A common objection to this is that contemporary hunter–gatherers show variable residence patterns, even cases in which men preferentially reside with their wife's kin.

This for Chapais is irrelevant, as his argument is about origins, not contemporary adaptations to socioecological constraints. Many will contest the assumption of male philopatry, especially adherents to the idea that humans evolved as cooperative breeders. These scholars argue that women's extended post-reproductive lifespans, short intervals between births and the extreme and extended helplessness of our children evolved because of the aid of maternal grandmothers and other kin in rearing children.

Other contentious claims include the irrelevance of infanticide and parental investment in the evolution of pair bonds. The book's treatment of proposed alternatives for pair-bond origins — a pact among individuals to reduce the costs of a physical scramble for mates — is dissatisfying because Chapais fails to consider sexual size dimorphism, brain expansion and changes in life-history traits across the paleoanthropological record. He also argues that food sharing arose from bipedalism, not from male specialization in hunting. Moreover he dismisses language as an important step in our becoming human, a position that will alienate many social scientists. The answer to the perennial question — why did this suite of traits arise only in hominids — is not fully dealt with, despite deft intellectual fencing.

Refreshingly, Chapais does not seek specific selective pressures for every trait. He conceives of evolution occurring under the constraints of prior adaptations and producing novel features from pre-existing parts. He abhors unsubstantiated evolutionary narratives and nimbly marshals evidence from primatological studies to sociocultural analyses to support his case. In the end, his book offers us one more scenario of our human trajectory, but it is a scholarly one. Chapais' thesis urges us to consider very carefully why humans are so different. ■

Monique Borgerhoff Mulder is professor at the Department of Anthropology, Graduate Group in Ecology, and Center for Population Biology at the University of California, Davis, 1 Shields Avenue, Davis, CA 95616. She is co-author of *Conservation: Linking Ecology, Economics and Culture*.

Novel alchemy

The Sun and Moon Corrupted
by Philip Ball
Portobello Books: 2008. 423 pp. £10.99

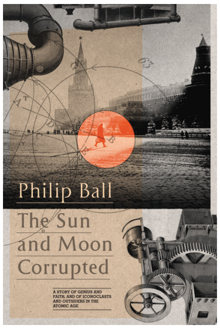
How do you write a novel about science? Many people do it badly, but with *The Sun and Moon Corrupted*, Philip Ball succeeds at his first attempt. Six steps lead to a good read, and Ball has aced them all (almost).

First, the book must have some science in it. Every tribe likes to see its rituals described. Ball relays the over-hearty greetings that anxious newcomers call out at scientific conferences. He mentions the nutty letters that serious researchers get from 'mad' outsiders: "To my last letter I have not received an answer ... You have to send all three papers for composition ... Any day of delaying its announcement costs milliards of dollars."

Second, it needs more than science. Recounting surface traits is fun, but on its own produces a standard academic novel in which nothing of depth lingers after the reader's delight at self-recognition has passed. A lesson in abject failure is provided by Tom Stoppard. In his play *Hapgood*, Stoppard pairs ambiguities in quantum mechanics with those in the British spy world. The play fails because there is no inherent link between quantum quirks and secret intelligence. Stoppard could have stripped the science without affecting the plot.

Ball doesn't throw in his science as an optional extra — it infuses his descriptions. Where laundry blows in the wind on the balconies of abandoned high-rise blocks, he notes with a chemist's eye that "already the dyes were burning, breaking up, and fading, the brightest and cheapest first, in the sun's strong glare." On a deeper level, his plot — of how a once-sensible researcher becomes a 'nutty' outsider who believes in perpetual motion — depends on science for its essence.

The third step is not to forget people. Without believable characters you have not a novel but a disquisition, chunkily hidden under dialogue. Bertrand Russell was an eloquent writer of non-fiction and an important logician, but his fiction was as convincing as Rupert Murdoch talking about ethics. Ball is far better. Karel Neder, the researcher he tracks in *The Sun and Moon Corrupted*, is believable. We meet him in his early teens as he discovers the beauties of science. He thrills at finding the first friends he can share his excitement with at university; and realizes that although he's a good student, others are better. He wonders how he will ever compete with them in a research career.



Step four, weave a story. Because science is taught as a sequence of inevitable breakthroughs, science novelists often copy that structure. Indirect approaches are more compelling. Ball starts his novel with a corker of a mystery. An intense young woman, Lena, walks in an abandoned city. She eyes the laundry still hanging, and observes rows of shoes in the town's kindergarten, all neatly in order, with a name painted inside each one. We wonder why the town has been left, and why she is there.

Step five is to be rational. It's a treat to follow Lena's quest. She's a new journalist, trying to make a living as a freelancer, but her heart is not in it. More important is her relationship with her father, a physicist at a British university. She can't bear to disappoint him, yet his smug rationality gives her little space to breathe. When Lena hears of the now elderly Neder's work, she realizes that his talk of perpetual motion machines is nonsense, and wants to know how he parted from the scientific mainstream. She reconstructs his path: from his escape from his native Hungary to success in the United States, before returning to Europe, where he exists at the fringes of academia. The levels in Ball's plot hold together like a musical chord. Will understanding Neder's life allow Lena to understand both her father and herself?

The sixth lesson is to avoid being too rational. Ball's writing is hyper-realistic, and he gives accounts of scientific conferences, journalists and secret police, yet still conveys a feeling that something else is going on, something important that lurks below our consciousness. The abandoned city is permeated with the mood from the Bible's book of Revelation: "And the city had no need of the sun, neither of the moon, to shine in it." As the narrator talks about dye sequences and Lena tries to grasp what drove Neder, there are deeper, almost alchemical, forces at work. ■

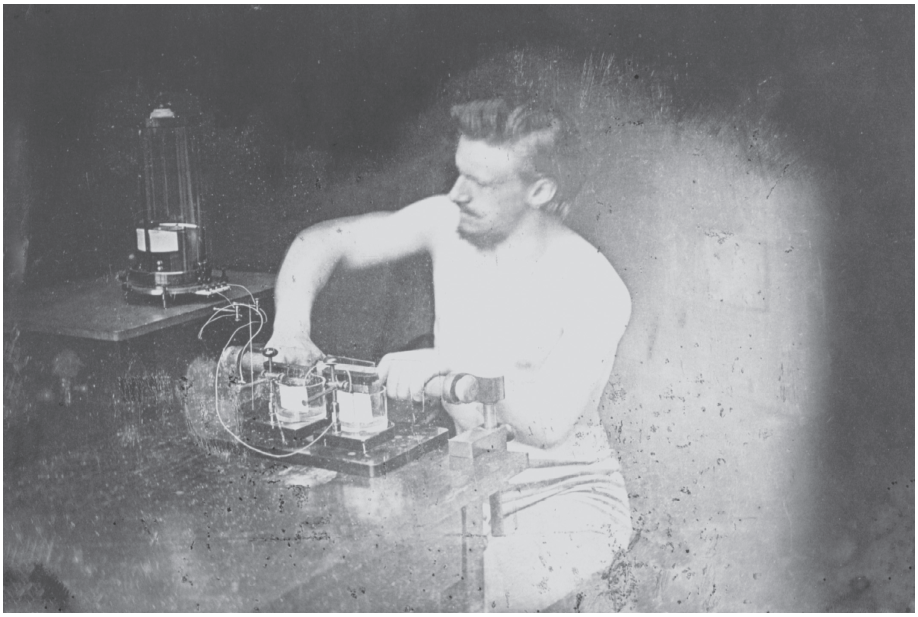
David Bodanis is the author of *Passionate Minds: The Great Scientific Affair*.

Hidden Treasures: Institute of Physiology collection

A cache of beautiful nineteenth-century German scientific devices that has survived many misfortunes now needs a new home, reports **Alison Abbott**.

The ingenious 1847 'Multiplikator' was the finest instrument of Emil du Bois-Reymond (1818–96), the first scientist to measure current flowing in nerves and muscles in the working human body. Built in Germany, in collaboration with instrument-maker Johann Georg Halske, the device's sleek, high-precision form presages the Bauhaus aesthetic of eight decades later. With the Multiplikator, du Bois-Reymond consolidated his position as a founder of modern electrophysiology.

The 1840s launched electrophysiology — some 75 years after Luigi Galvani made dead frog muscle twitch by applying electric current to a nerve. For most of his life Galvani fought to persuade sceptical colleagues that the apparent quackery of 'animal magnetism' and 'animal electricity' was a reality. He was long dead before the instrument bearing his name, the galvanometer, was invented. And it was du Bois-Reymond who rendered the galvanometer sensitive enough to detect the tiny physiological currents in living tissues; the currents induce magnetic waves that deflect a compass needle.



BPK / BERLIN STATE LIBRARY - PRUSSIAN CULTURAL HERITAGE, MANUSCRIPT DEPARTMENT

The Multiplikator, one of the fine examples of instrument craftsmanship in the Johannes Müller collection.

Around the Multiplikator's core, du Bois-Reymond wound more than 10,000 coils of thin wire to magnify the effect of the current. He connected the wires' ends to two vessels of salt solution. The Multiplikator operator — usually du Bois-Reymond himself — held two wooden batons while dipping one finger of each hand into the salt baths to close the circuit. He clenched one arm tightly, keeping the other relaxed. The transient negative electrical potential induced around the tensed muscle caused a current to flow between his arms, and through the entire circuit. The current's strength was revealed by the angle of the needle's deflection. To reduce resistance in his body, du Bois-Reymond worked bare-chested. To generate the strongest possible currents, the former competitive gymnast trained obsessively in a gym he built in his home.

It was a time when scientists took aesthetics seriously. du Bois-Reymond photographed his brother operating the Multiplikator (pictured), then turned the image into an engraving of Apollo, the Greek god of the arts, who embodies perfect male beauty.

Berlin was a major centre for science in the nineteenth century. Many of its most successful researchers were students at the Institute of Anatomy at the Friedrich Wilhelm University (now Humboldt University). These included du Bois-Reymond as well as the physicist and physiologist Hermann von Helmholtz, the zoologist and artist Ernst Haeckel, and Werner von Siemens. Siemens helped to develop the Multiplikator before forming a telegraphic company in 1847, the forerunner of today's Siemens electronics behemoth. He took Halske with him, a blow to the many Berlin scientists for whom the artisan had crafted unique physiological instruments.

By the late 1840s, the institute had a major collection of instruments, and it added more over the next few decades. When du Bois-Reymond was made director of a new, purpose-built Institute of Physiology in 1877, he displayed many of the instruments in its foyer. A year later, some of them were destroyed in a fire.

That was just the start of the collection's misfortunes. Having survived the Nazi period, the institute found itself in the Russian sector of occupied Berlin in 1945. Over the next 50 years the increasingly decrepit and destitute organization had more than 20 directors.

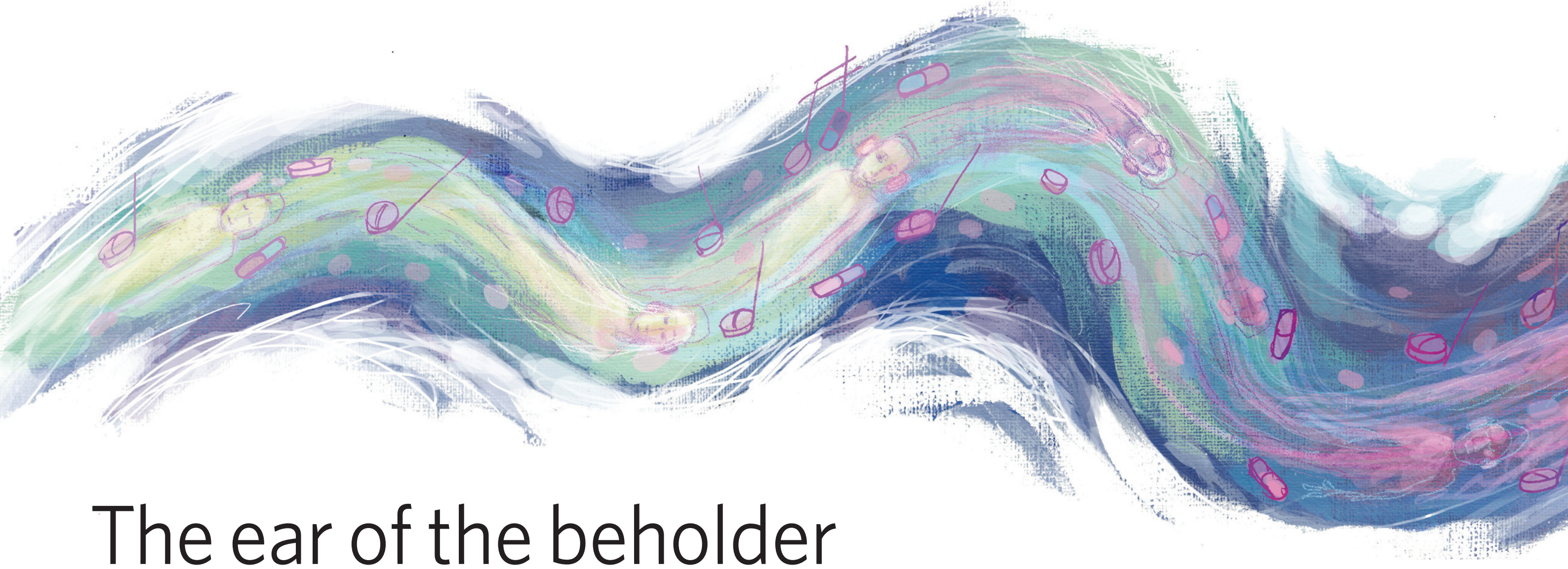
The historical instruments were regularly plundered for parts to build new gadgets for sexier sciences, notably biochemistry. When the institute's lecture theatre was reconstructed in the late 1950s, many of the glass cabinets in which the collection had been stored disappeared, to be rediscovered when the building's attic was cleared out in 1975. Those instruments that were not beyond rescue were cleaned and renovated, and displayed for the first time during the celebrations of the institute's centenary, in 1977.

After Berlin was reunified in 1990, the university physiology department had to move again — to a building that is now being sold by the cash-strapped organization Charité, a recent amalgamation of Berlin's eastern and western medical university faculties. Despite its chequered past, the instrument collection still comprises more than 200 important objects, including Helmholz's beautiful series of brass resonators and other devices built for his seminal work on physiological acoustics. Now named after the great experimental anatomist Johannes Müller, it also boasts many switching contraptions that delivered controlled pulses of electricity to the biological preparations and chymographs that recorded contractions of prepared muscles by scratching the smoked surface of a slowly revolving drum. First though, the collection needs to secure a new home. ■

Alison Abbott is *Nature*'s senior European correspondent.

See www.charite.de/ch/physio/hi/start.html for details. For more Hidden Treasures see www.nature.com/nature/focus/hiddentreasures.

ESSAY



The ear of the beholder

In the last of nine Essays on science and music, **John Sloboda** argues that researchers must study music as people actually experience it, if they are to understand how it affects thoughts and feelings.

Thanks to technology, we have an unprecedented choice of music to listen to, and places and times to hear it. Music has never been more accessible. But never has it been more mysterious and inscrutable. Many people say they lack musical skills such as singing, playing an instrument or composing. A fifth of adults believe they are 'tone deaf', so they don't see music as something they do; rather, they experience music as something that is done to them, something that, at a very deep level, they don't fully understand.

This contemporary disconnection may be limiting the scientific study of music. If research is to provide a satisfying account of how music and mind interact, it must embrace the full variety of musical experiences and contexts. Music science is now mature enough to take more risks in the scope of its investigations. If science is to demystify music and explain its power to affect us, it must investigate music as it is actually experienced.

The vitamin model

The musically disempowered view music as something that is provided for 'consumers' or 'patients' by a range of professionals: composers, producers, performers, even health workers. People have lost confidence in their own judgements and intuitions about it.

This disconnection has really only occurred in the past 50 years and is most common in prosperous industrialized nations. In other times and places, people have had a much deeper relationship with music, because they helped to make most of the music in their lives, through community activities involving singing and dancing, and the passing down of an oral tradition. The modern perplexity with music may be a symptom of the loss of this natural connection to making music. People can surround themselves with highly polished professional performances at the press of a button, so there is less incentive to sing or play.

This cultural change means that the musical repertoire is seen as a box of pills, with different pieces having different active properties, such as sedation or aphrodisia. This 'vitamin model' lies behind misleading commercial initiatives that package selections of music as stress busters, relaxation aids or even music to make your baby more intelligent.

But music cannot achieve a prescribed psychological outcome because it is impossible to understand or predict its effects without accounting for two other significant areas of influence. The first relates to the listener: his moods, memories, intentions, attitudes, choices and experiences. The second is the social context in which the music is experienced: who else is there, what is going on, and the social or personal significance of the event.

Experimental science often inadvertently promotes the vitamin model. Scientists who study music are not, as a body, skilled in musical performance or composition. Our relationship to music is, like much of society, generally that of an outsider looking in, a consumer rather than producer.

Additionally, research methods regularly minimize the role of the listener in the construction of a response. We give research participants little choice in how to respond, and we tend to draw our conclusions from sample averages rather than individual data. The role of social context is also reduced in the laboratory, where the subject is removed from the usual situations where they would engage with music. Such simplification may be needed to kick-start early explorations of a complex topic, but research must not get stuck at this level. It needs to encompass the experienced reality of the phenomenon.

Consider two situations in which music is heard. In the first, an experimental subject lies with her head clamped inside a functional magnetic resonance imaging chamber. Surrounded by strangers in white coats, she is played short pairs of synthesized chord sequences and asked whether the two sequences are identical or not. She finds it boring and rather uncomfortable.

In the second situation, a parent sits in a school hall with other parents, watching her six-year-old daughter give her first public musical performance. She is totally absorbed in the event. Her heart swells with tearful pride


D. PARKINS

and relief as her daughter plays Pachelbel's *Canon* without serious mishap.

Context is king

Even if the same musical stimulus were heard in both situations, the lab study would provide little insight into the school-concert situation. This inability to make direct mappings from a musical stimulus to a predictable psychological effect is not just an inconvenient difficulty. Rather, it points to a profound reality: music is a constantly evolving product of human culture, whose forms, functions and discourses are not fixed. It is what we decide to make it. Music originally written as a profound statement of religious belief can be used as background music for domestic chores, or as a stimulus for a laboratory experiment. And its effects in these different contexts can be quite different.

We must also avoid the trap of asserting that music is so complex and ineffable that a detailed scientific understanding of it is impossible, and that it must remain shrouded in mystery. On this issue, musicians and scientists often line up on opposite sides, with musicians seeing scientists as naive over-simplifiers, and scientists regarding musicians as defensive 'obscurantists'.

Some scientific research has already begun to provide systematic accounts of why our response to music is so variable. Our responses are hugely influenced by our attitudes to particular musical styles, and whether we approve of or identify with them. They are equally influenced by our sense of whether the music is appropriate to the situation. Some people find music in shops and restaurants aggravating; others find its absence disconcerting.

Research that takes us beyond the vitamin model is likely to have one or more of the following characteristics. It studies music in real-life, non-laboratory settings. It studies music across a range of contexts, from music in public places that is barely attended to, through to deliberately chosen music that is given full attention. It studies the nature of musical choices that individuals make. It uses a combination of observation, experience-sampling and post-event interviewing to draw out the personal and social meanings of the musical events for the listener. It uses complete, authentic musical objects, such as whole songs, rather than impoverished or machine-constructed stimulus segments. Finally, it studies effects of musical stimuli that are time sensitive and linked to the detailed unfolding pattern of events, rather than studying unspecific effects of entire pieces. The first few seconds of a piece may fill a listener with excited expectation, and the end with a sense of relief or homecoming. You cannot get at this from a global measure of how 'happy' a piece is.

A growing body of research is exploring the richness of the human musical experience without trying to control or limit it. For example, Tia De Nora at the University of Exeter has carried out pioneering research through in-depth interviews with women about their complex experiences of music in such contexts as shops, gyms and the home. Key to these studies was her extensive probing of the different factors at work in each situation, and the self-conscious decisions about what to listen to, and how to listen to it.

Building on this work, several investigators, including my group, have started charting individual musical experiences over extended time periods. By having frequent telephone conversations with participants over about a week, say, we have shown that musically untrained individuals interact with music a dozen times a day or more with varying attention and engagement. One important result is that the people we studied mainly use music as an accompaniment or enhancement to some other activity, and use it for a variety of functions, including to regulate their mood or attention, and as a signal of cultural identity.

Listen and learn

Such studies point to a way forward for music science that takes seriously listeners' beliefs, feelings and situations.

One beneficial effect of the careful scientific probing of listeners' experiences is that it often demonstrates their hidden musical competence. Studies of encoding and memory reveal musical intelligence in people's recall errors: they tend to substitute a note or chord that serves a similar musical function. This shows that they have subconsciously internalized the rules of musical grammar. Other studies show that the ability to sing in tune can be dramatically improved by simple well-targeted feedback, suggesting that many abilities are already in place but are masked by the absence of one simple cognitive component.

Experiments also show that listeners can acquire a complex and fine-grained appreciation of the link between the choices they make about what to listen to and the resultant psychological outcomes. They come to know which pieces of music reliably uplift or calm them, or remind them of a valued truth. This ability, even if idiosyncratic and unschooled, is the foundation of the nuanced internal representations of music that are prerequisite for the development of skilled performance.

Careful scientific research is beginning to chart the complex nature of everyday musical expertise. It shows that music is neither an unfathomable mystery nor a magic pill, but a human construction that can be used self-consciously and skilfully by anyone for a variety of cognitive and emotional purposes, mundane and profound. ■

John Sloboda is professor of psychology and an honorary research fellow at the Institute of Law, Politics and Justice, Keele University, Newcastle, Staffordshire ST5 5BG, UK. He is the author of *Psychology for Musicians* (Oxford University Press, 2007).

For further reading, see http://tinyurl.com/6mkupl. To read all nine Science & Music Essays, see www.nature.com/nature/focus/scienceandmusic.

ESSAY

Arise 'cliodynamics'

If we are to learn how to develop a healthy society, we must transform history into an analytical, predictive science, argues **Peter Turchin**. He has identified intriguing patterns across vastly different times and places.

What caused the collapse of the Roman Empire? More than 200 explanations have been proposed[1], but there is no consensus about which explanations are plausible and which should be rejected. This situation is as risible as if, in physics, phlogiston theory and thermodynamics coexisted on equal terms.

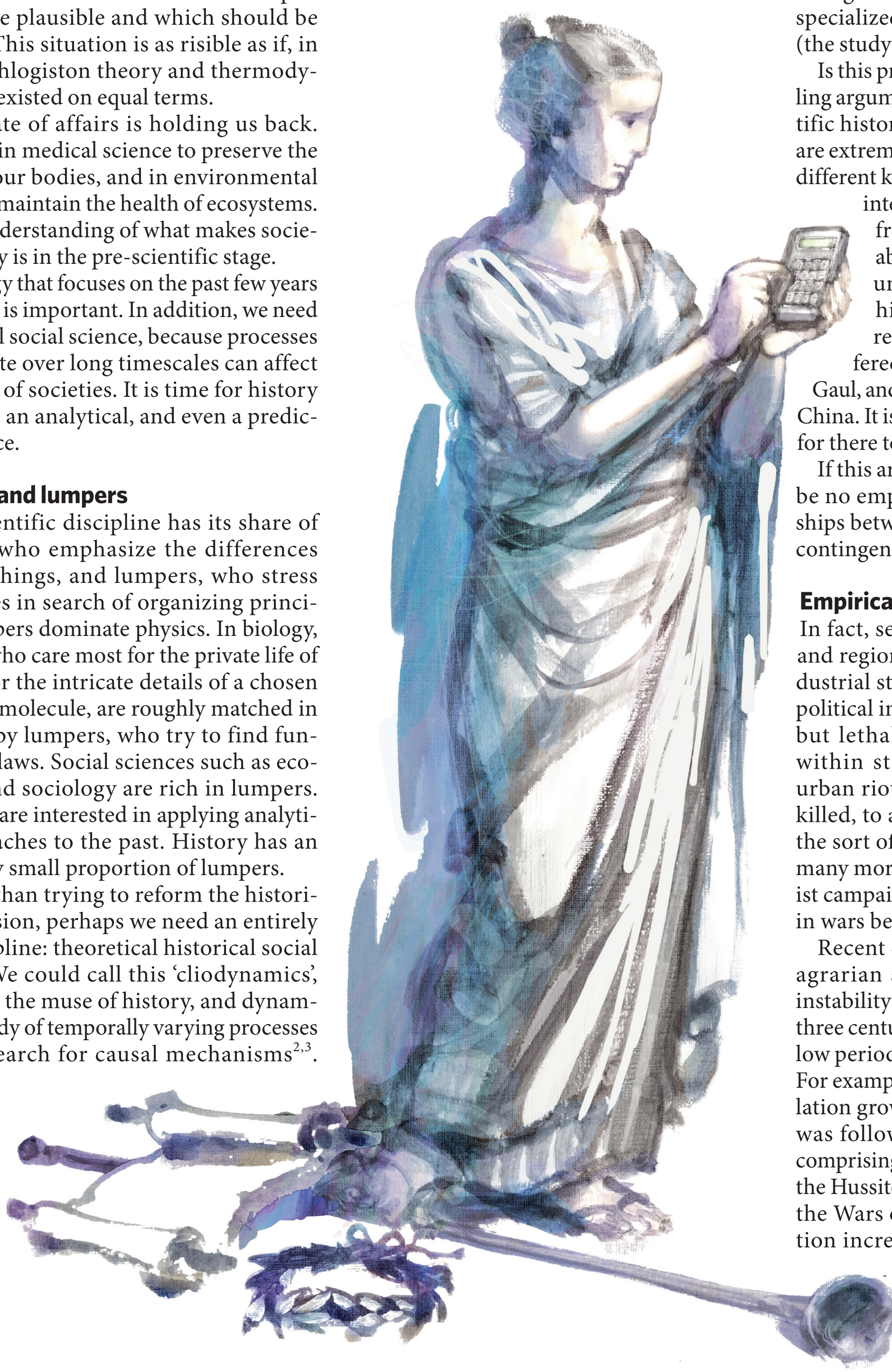
This state of affairs is holding us back. We invest in medical science to preserve the health of our bodies, and in environmental science to maintain the health of ecosystems. Yet our understanding of what makes societies healthy is in the pre-scientific stage.

Sociology that focuses on the past few years or decades is important. In addition, we need a historical social science, because processes that operate over long timescales can affect the health of societies. It is time for history to become an analytical, and even a predictive, science.

Splitters and lumpers

Every scientific discipline has its share of splitters, who emphasize the differences between things, and lumpers, who stress similarities in search of organizing principles. Lumpers dominate physics. In biology, splitters, who care most for the private life of warblers or the intricate details of a chosen signalling molecule, are roughly matched in numbers by lumpers, who try to find fundamental laws. Social sciences such as economics and sociology are rich in lumpers. Sadly, few are interested in applying analytical approaches to the past. History has an alarmingly small proportion of lumpers.

Rather than trying to reform the historical profession, perhaps we need an entirely new discipline: theoretical historical social science. We could call this 'cliodynamics', from Clio, the muse of history, and dynamics, the study of temporally varying processes and the search for causal mechanisms[2,3].



D. PARKINS

Let history continue to focus on the particular. Cliodynamics, meanwhile, will develop unifying theories and test them with data generated by history, archaeology and specialized disciplines such as numismatics (the study of ancient coins).

Is this proposal feasible? The most compelling argument against the possibility of scientific history goes like this. Human societies are extremely complex. They consist of many different kinds of individuals and groups that interact in complex ways. People have free will and are therefore unpredictable. Moreover, the mechanisms that underlie social dynamics vary with historical period and geographical region. Medieval France clearly differed in significant ways from Roman Gaul, and both were very different to ancient China. It is all too messy, argue the naysayers, for there to be a unifying theory.

If this argument were correct, there would be no empirical regularities. Any relationships between important variables would be contingent on time, space and culture.

Empirical empires

In fact, several patterns cut across periods and regions[3]. For example, agrarian, preindustrial states have seen recurrent waves of political instability — not interstate warfare, but lethal collective violence occurring within states, ranging from small-scale urban riots, in which just a few people are killed, to a full-blown civil war. This is just the sort of violence we need to understand: many more people are killed today in terrorist campaigns, civil wars and genocides than in wars between nations[4].

Recent comparative research shows that agrarian societies experience periods of instability about a century long every two or three centuries. These waves of instability follow periods of sustained population growth. For example, in Western Europe, rapid population growth during the thirteenth century was followed by the 'late-medieval crisis', comprising the Hundred Years War in France, the Hussite Wars in the German Empire, and the Wars of the Roses in England. Population increase in the sixteenth century was followed by the 'crisis of the seventeenth century' — the wars of religion and the Fronde in

France, the Thirty Years War in Germany, and the English Civil War and Glorious Revolution. Similarly, population growth during the eighteenth century was followed by the 'age of revolutions', ranging from the French Revolution of 1789 to the pan-European revolutions of 1848–49 (ref. 5).

Such oscillations between population growth and instability have been termed 'secular cycles'[6]. Given the limitations of historical data, we need an appropriately coarse-grained method to determine the statistical significance, and the generality, of the pattern. The basic idea is to demarcate population growth and decline phases, and to count the instability incidents (such as peasant uprisings and civil wars) that occur during each phase.

With my colleagues Sergey Nefedov and Andrey Korotayev, I have collected quantitative data on demographic, social and political variables for several historical societies. Applying the above approach to eight secular cycles in medieval and early modern England, France, the Roman Empire and Russia, we find that the number of instability events per decade is always several times higher when the population was declining than when it was increasing[6]. The probability of this happening by chance is vanishingly small. The same pattern holds for the eight dynasties that unified China, from the Western Han to the Qing[7], and for Egypt from the Hellenistic to the Ottoman periods[8].

"Societies as different as medieval France, the Roman Empire and China under the Han dynasty share dynamics."

Making waves

Such strong regularity points to the presence of some fundamental principles. Population growth beyond the means of subsistence leads to declining levels of consumption and popular discontent, but this is not enough to destabilize agrarian societies. Peasant uprisings have little chance of success when the governing élites are unified and the state is strong[9].

The connection between population dynamics and instability is indirect, mediated by the long-term effects of population growth on social structures. One effect is the increasing number of aspirants for élite positions, resulting in rivalry and factionalism. Another consequence is persistent inflation, which causes a decline in real revenues and a developing fiscal crisis of the state. As these trends intensify, they result in state bankruptcy and a loss of military control; conflict among élite factions; and a combination of élite-mobilized and popular uprisings, leading to the breakdown of central authority[3,9].

This explanation — the 'demographic–structural' theory — is a work-in-progress. Our tests with the eight case studies[6] support some of its predictions: for example, élite overproduction preceded the crisis in every case. The tests also identify areas where the theory needs to be modified. Perhaps we need an entirely new theory to explain the observed patterns and predict new ones, but that is the business of science. The important thing is that societies as different as medieval France, the Roman Empire and China under the Han dynasty share dynamics, when viewed in an appropriately coarse-grained way. Not everything in history is contingent and particular.

Even so, theories developed and tested on preindustrial data must be modified before they can be applied to contemporary social dynamics. Happily, there are indications that our theories will not need to be replaced wholesale. Rapid demographic change and élite overproduction were still important factors in twentieth-century revolutions[10].

Furthermore, over the past 200 years, political instability in the United States has waxed and waned in a pattern reminiscent of that in preindustrial societies. Political violence — urban riots, lynchings, violent labour disputes and so on — was almost absent in the early nineteenth century, increased from the 1830s and reached a peak in around 1900. The American Civil War occurred during this period of growing unrest. The instability then subsided during the 1930s, and the following two decades were remarkably calm. Finally, in the 1960s, political violence increased again[11].

It remains to be seen whether a modified version of the demographic–structural theory can explain this pattern. The point is that the study of such slow-moving processes requires a long-term view and an explicitly historical approach.

Learning lessons

Any claim that history can become a predictive science raises eyebrows. But scientific prediction is a broader concept than merely forecasting the future. It can be used to test theories. For example, two rival theories may make different predictions about the behaviour of some variable, such as birth rate, under certain social conditions. We then ask historians to explore the archives, or archaeologists to dig up data, and determine which theory's predictions best fit the data. Such retrospective prediction, or 'retrodiction', is the life-blood of historical disciplines such as astrophysics and evolutionary biology.

Cliodynamic theories will not be able to predict the future, even after they have passed empirical tests. Accurate forecasts are often impossible because of phenomena such as mathematical chaos, free will and the self-defeating prophecy. But we should be able to use theories in other, perhaps more helpful, ways: to calculate the consequences of our social choices, to encourage the development of social systems in desired directions, and to avoid unintended consequences.

Like other systems with nonlinear feedback, societies often respond to interventions in surprising ways. When the Assembly of Notables refused to approve a new land tax in 1787, they did not intend to start the French Revolution, in which many of them lost their heads. When Tony Blair was Britain's prime minister, he set out to increase the proportion of youth getting higher education to 50%. He was presumably unaware that the overabundance of young people with advanced education preceded the political crises of the age of revolutions in Western Europe[12], in late Tokugawa Japan and in modern Iran and the Soviet Union[9,10].

It is time we heeded the old adage that those who do not learn from history are doomed to repeat it. We must collect quantitative data, construct general explanations and test them empirically on all the data, rather than on instances carefully selected to prove our pet narratives. To truly learn from history, we must transform it into a science. ■

Peter Turchin is professor of ecology and mathematics at the University of Connecticut, Storrs, Connecticut 06269, USA. He is the author of *War and Peace and War: The Life-Cycles of Imperial Nations* (Pi Press, 2006).

1. Demandt, A. *Der Fall Roms: die Auflösung des Römischen Reiches im Urteil der Nachwelt* (Beck, Munich, 1984).
2. Turchin, P. *Historical Dynamics: Why States Rise and Fall* (Princeton Univ. Press, 2003).
3. Turchin, P. *War and Peace and War: The Life Cycles of Imperial Nations* (Pi Press, 2006).
4. Mack, A. (ed.) *Human Security Report 2005: War and Peace in the 21st Century* (Oxford Univ. Press, 2005).
5. Fischer, D. H. The Great Wave: Price Revolutions and the Rhythm of History (Oxford Univ. Press, 1996).
6. Turchin, P. & Nefedov, S. *Secular Cycles* (Princeton Univ. Press, 2008).
7. Nefedov, S. PhD dissertation [in Russian] (Ekaterinburg Univ., 1999).
8. Korotayev, A. & Khaltourina, D. *Introduction to Social Macrodynamics: Secular Cycles and Millennial Trends in Africa* (URSS, 2006).
9. Goldstone, J. A. *Revolution and Rebellion in the Early Modern World* (Univ. California Press, 1991).
10. Goldstone, J. A. *J. Int. Affairs* **56,** 3–21 (2002).
11. Levy, S. G. *Political Violence in the United States, 1819-1968.* (Computer file, Inter-University Consortium for Political and Social Research, Ann Arbor, 1991).
12. O'Boyle, L. *J. Modern Hist.* **42,** 471–495 (1970).

NEWS & VIEWS

ORIGINS OF LIFE

How leaky were primitive cells?

David W. Deamer

If the first cells were simple vesicles, how did nutrients cross their membranes without help from transport proteins? A model of a primitive cell suggests that early membranes were surprisingly permeable.

How life began remains an open question. There are now a dozen or more competing ideas that fall into two general categories: life began either as an autotrophic organism that used primitive metabolic pathways to make its own organic components, or as a heterotroph that incorporated carbon-containing nutrients already available in the environment. On page 122 of this issue, Mansy *et al.*[1] weigh in with a laboratory simulation that supports a heterotrophic origin of life — a cell-like vesicle that allows small, organic 'nutrient' molecules to pass through its membrane.

Recent progress in planetary science has provided a reasonably convincing description of what our planet was like at the time of life's origin, sometime between 3 billion and 4 billion years ago. It was a world of global oceans peppered with volcanic land masses resembling the relatively newborn lands of today's Hawaii and Iceland. The atmosphere was a mixture of carbon dioxide and nitrogen, with little or no oxygen, and the mean global temperature was much higher than that today, probably in the range 60–70 °C. Seas and lakes contained dilute mixtures of simple organic compounds, perhaps with thin films of oily hydrocarbons on the surface. But most importantly for a discussion of Mansy and colleagues' work[1], local conditions were far from equilibrium — a constant flux of energy drove organic reactions towards ever-increasing complexity. This would ultimately have yielded various polymeric products, perhaps including prototypes of nucleic acids or proteins.

The result of this process would have been that vast numbers of microscopic assemblies of molecules (Fig. 1) appeared wherever organic compounds became concentrated at the interface between the atmosphere, water and mineral surfaces. In one scenario that might have led to the development of heterotrophic life, these assemblies took on a cell-like form — membranous compartments composed of bilayers, within which molecules are trapped. Such compartments are known to readily self-assemble from amphiphilic compounds (soap-like fatty acids and fatty alcohols) that are commonly found in experiments that simulate the prebiotic 'soup'.

A. NORTH, ROCKEFELLER UNIV.

Figure 1 | Lipid self-assembly. A mass of phospholipid molecules (the building-blocks of cell membranes) extends tubular structures as it absorbs water, in a classic example of the self-assembly processes that might have led to the spontaneous formation of cell-like structures on prebiotic Earth. Mansy *et al.*[1] show that a protocell-like vesicle that self-assembles from simple organic molecules in a similar way doesn't need proteins to transport molecules across its membrane.

In this theory of the origins of life, each cell-like assembly had a different composition from the next. Most were inert, but a few might have contained a particular mixture of components that could be driven towards further complexity by capturing energy and small 'nutrient' molecules from the environment — the beginnings of a heterotrophic system. As the nutrient molecules were transported into the internal compartment, they became linked together into long chains in an energy-consuming process. Life began when one or more of the assemblies found a way not only to grow, but also to reproduce by incorporating a cycle involving catalytic functions and genetic information.

So far, so plausible. But there's a problem: the membrane that forms the compartment of the putative cell is also a permeability barrier. Small molecules that have no electrical charge (such as water and carbon dioxide) get across in seconds, but larger molecules (such as amino acids) that have electrically charged groups are a billion times less permeable. Modern cells solve this problem by integrating proteins into their membranes, with each protein specialized for transporting specific molecules such as amino acids, glucose or phosphate. But how did the first forms of cellular life overcome the permeability problem?

This is the question posed by Mansy *et al.*[1]. Specifically, they ask how growing systems of polymers in primitive cells could have had access to nutrients in the external environment. To find the answer, the authors made membrane-bound compartments not just from modern phospholipids (the building-blocks of cells that have evolved to be virtually impermeable to ionic solutes), but also from mixtures of simpler molecules, such as fatty acids, fatty alcohols and monoglycerides (which might have been present on prebiotic Earth). Earlier work demonstrated that such mixtures produce surprisingly stable membranous vesicles[2,3].

Mansy and colleagues first investigated the permeability of the vesicles to ribose, the sugar component of RNA. They optimized the mixture of molecules in the vesicle membranes for maximal permeability to ribose, but minimal permeability to polymers such as DNA. In this way, the authors showed that membranes similar to those that might have formed on prebiotic Earth allow the passage of simple molecules, as would be essential for a functioning protocell.

They then trapped synthetic DNA molecules in the optimized vesicles. The DNA was designed to act as both a primer and a template for its own elongation, so that the primer spontaneously extends in the presence of the appropriate molecular building-blocks — nucleotides that are chemically 'activated' to undergo polymerization on the template. The template was a single strand of DNA that contained only cytosine bases, so the activated nucleotides had

to contain the complementary guanosine bases. Mansy *et al*. added these nucleotides to the medium surrounding the vesicles, where they are analogous to the external nutrients required by a heterotrophic organism.

For vesicles composed of phospholipids, the authors observed no elongation of the encapsulated DNA primer on addition of guanosine-containing nucleotides to the external medium. This was to be expected, because the nucleotides were excluded from the interior volume. But the vesicles composed of simple amphiphilic molecules showed a remarkable elongation of the DNA primer over the course of 24 hours, as nucleotides were added one by one. This demonstrates that a primitive cellular compartment can be permeable enough to let a cell grow by allowing small molecules through the bilayer, yet stable enough to maintain the integrity of the internal mixture of functional polymers. Previous work[4–7] has shown that a variety of polymerization reactions work beautifully when encapsulated in different systems, but Mansy *et al.*[1] establish for the first time that a simulated prebiotic protocell can work with an external source of reagents.

Arguably the most important aspect of this study is that it demonstrates that a heterotrophic origin of primitive cellular life is feasible. The next step is to devise some version of non-enzymatic replication in which a polymer such as DNA or RNA is completely reproduced within a vesicle using externally added substrates. That's a tall order, of course, but presumably not an impossible one — at some point in the pathway to the first cellular life, such a system must have arisen spontaneously.

Cells are the basic unit of all life today, and there is increasing reason to think that the first form of life was a primitive version of a cell, rather than a replicating molecule supported by a metabolic network[8]. Mansy and colleagues' work[1] suggests that the first cellular life might have sourced energy and nutrients from the environment, and that more complex, autotrophic lifestyles appeared at a later stage of evolution. We may never know for sure if this is true, but we certainly now have some compelling circumstantial evidence. ■

David W. Deamer is in the Department of Biomolecular Engineering, University of California, Santa Cruz, Santa Cruz, California 95064, USA.
e-mail: deamer@soe.ucsc.edu

1. Mansy, S. S. *et al. Nature* **454,** 122–125 (2008).
2. Apel, C. L., Deamer, D. W. & Mautner, M. N. *Biochim. Biophys. Acta Biomembranes* **1559,** 1–9 (2002).
3. Monnard, P.-A., Apel, C. L., Kanavarioti, A. & Deamer, D. W. *Astrobiology* **2,** 139–152 (2002).
4. Walde, P., Goto, A., Monnard, P.-A., Wessicken, M. & Luisi, P. L. *J. Am. Chem. Soc.* **116,** 7541–7547 (1994).
5. Noireaux, V. & Libchaber, A. *Proc. Natl Acad. Sci. USA* **101,** 17669–17674 (2004).
6. Ishikawa, K., Sato, K., Shima, Y., Urabe, I. & Yomo, T. *FEBS Lett.* **576,** 387–390 (2004).
7. Monnard, P.-A., Luptak, A. & Deamer, D. W. *Phil. Trans. R. Soc. Lond. B* **362,** 1741–1750 (2007).
8. Szostak, J. W., Bartel, D. P. & Luisi, P. L. *Nature* **409,** 387–390 (2001).

SOLAR SYSTEM

A shock for Voyager 2

J. R. Jokipii

The Voyager 2 spacecraft has now followed Voyager 1 into the region beyond the end of the supersonic solar wind, where the influence of interstellar space is growing — so opening a new age of exploration.

Five papers[1–5] in this issue, beginning on page 63, celebrate and record a notable event in the storied history of the Voyager space programme. Thirty years after its launch on 20 August 1977, Voyager 2 arrived at the 'heliospheric termination shock', adding important new data about this turbulent boundary in the outer Solar System to those collected by Voyager 1 four years ago[6–8].

For context, some Solar System geography is necessary. The Sun emits a supersonic flow of plasma (the solar wind) in all directions, carving out a vast spheroidal cavity in the interstellar plasma. This cavity is called the heliosphere, and is some 100–150 astronomical units (AU) in extent — an AU is the average distance from Earth to the Sun, about 150 million kilometres. The size of the heliosphere is determined principally by the properties of the solar atmosphere and the surrounding interstellar plasma, magnetic field and neutral particles. At large distances from the Sun, the solar plasma is so extended that it can no longer push back the interstellar plasma. Its supersonic flow, relative to the interstellar medium, ends at a spheroidal shock wave, the termination shock, where the wind abruptly slows. The region of slower flow beyond the shock is the heliosheath. Here, the interaction of the solar plasma with the interstellar gas becomes dominant. The heliosheath ends at the heliopause, beyond which is the interstellar plasma. This entire system can be recreated in a kitchen sink, as shown in Figure 1.

Beginning on 31 August 2007 and ending the next day, Voyager 2 crossed the heliospheric termination shock several times, a feat made possible because this boundary is far from static and oscillates in and out. So the spacecraft has left the supersonic solar wind behind, and follows Voyager 1 into the heliosheath and on to the heliopause where they will probably arrive in the next decade or so (the distance to the heliopause is not yet known with any certainty, but is expected to be at about 130–150 AU from the Sun). The plasma analyser on Voyager 1 has long been inoperative. But that on Voyager 2 is working, and both spacecraft can measure plasma waves,

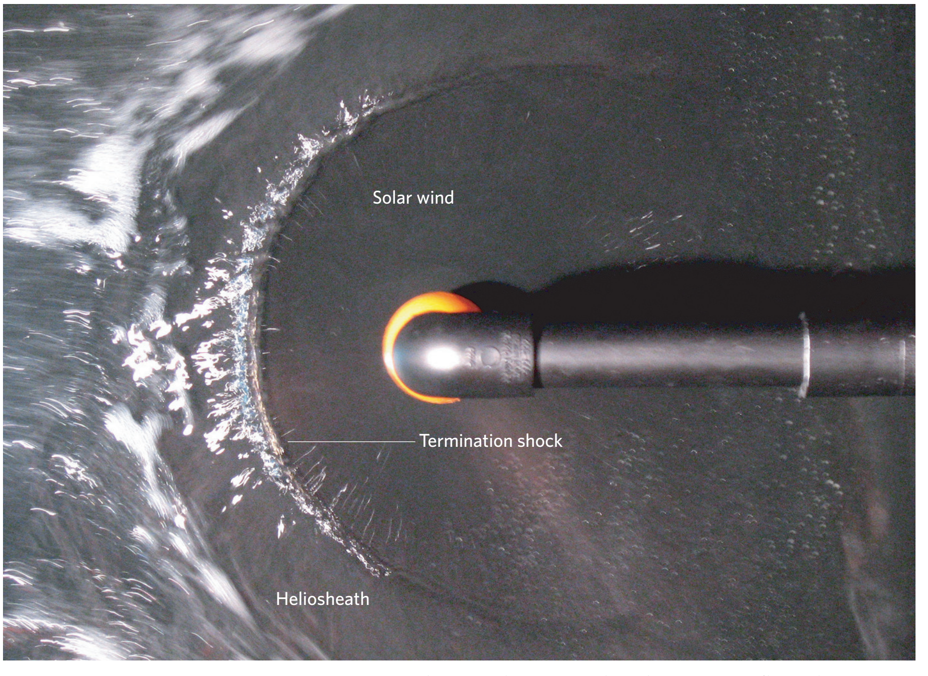


Figure 1 | The heliosphere in a kitchen sink. In this two-dimensional analogue, water flows down into the plane of the photograph from a tap (orange) and strikes a flat, inclined surface. The water radiates from the point of impact, much as the solar wind flows outwards from the Sun. It is eventually checked by surrounding water and forms a jump (a hydraulic bore, analogous to the termination shock) that is quasi-circular. The flow of the interstellar medium is represented by a uniform, slow flow of soapy water in from the left. The jump in this analogue is turbulent and dynamic, much as the termination shock is observed to be. (Analogue created by K. C. Hsieh, Department of Physics, University of Arizona. Photo by J. R. Jokipii.)

magnetic fields and energetic particles. At the heliopause, they will take *in situ* measurements of the interstellar plasma for the first time.

The first crossing of the termination shock, by Voyager 1, occurred when it was 94 AU from the Sun, during a several-month period of rapid (~100 km s^{-1}) inward motion of the shock. A gap in data transmission during this time meant that its detailed structure could not be determined. This first crossing nonetheless established a number of the shock's properties, including its distance, and the facts that it fluctuates considerably, moving in and out for months at a time, and that its overall shape is blunt[9] rather than bullet-shaped or rounded.

The crossings of the shock (at 84 AU) by Voyager 2 add to our understanding in several ways, as detailed in the five new papers[1–5], not least because the working plasma detector permitted a much richer set of observations. The fact that the Voyager 2 crossings occurred some 10 AU closer to the Sun than those of Voyager 1 confirms earlier suggestions[10] that the shock is laterally asymmetric (pushed in on one side), possibly because of the presence of an inclined local interstellar magnetic field[11–13]. The multiple crossings by Voyager 2 were presumably the result of dynamic changes in the shock — quick inward and outward motions, ripples propagating along the shock face or perhaps re-forming of the shock, phenomena much like those seen in the water analogue depicted in Figure 1.

The plasma data returned by Voyager 2 have also established the importance of interstellar 'pickup ions'. These are interstellar neutral atoms that have entered the solar wind and become ionized; they have a higher energy than the thermal particles, and strongly affect the processes at the shock. For a 'standard' shock, the downstream flow would be subsonic, but Voyager 2 found it to be cooler than expected and supersonic. Consistent with this, the higher-energy accelerated particles have an energy density comparable to that of the plasma and magnetic field. These facts establish that the shock at the Voyager 2 crossing is quite different from previously observed shocks, possibly because of the pickup ions.

The observations of the two Voyagers show that the termination shock is very complex, and many ambiguities remain. It will be a long time before we receive more *in situ* data, but — fortunately — remote observations from the inner heliosphere should fill some of the gaps. Electromagnetic waves from the solar-wind plasma in the heliosheath are too weak to be observable, but its interactions with ambient neutral atoms in the heliosheath give rise to energetic neutral atoms (ENAs) that can be observed at Earth (much as light or other radiation can). The IBEX spacecraft, to be launched later in 2008, will exploit this principle to provide a global view of the heliosheath. The recently launched STEREO spacecraft can also observe ENAs.

A paper by Wang *et al.*[14], also in this issue (page 81), reports on initial observations of these ENAs from the STEREO spacecraft, which complement the *in situ* Voyager observations from the heliosheath. The energy spectra are consistent with the Voyager observations, but the observed variation of the intensity with direction is unexpected. Particularly interesting and puzzling to this writer are the observed double maxima of the fluxes, as a function of longitude, near the direction from which the interstellar medium is flowing.

Over the past few years, the stream of *in situ* and remote data from the outer reaches of the heliosphere has revolutionized our view of how the Sun interacts with the Galaxy. More is to come as the Voyagers continue their journey, and as remote data, such as those from STEREO and the forthcoming IBEX mission, become available. ■

J. R. Jokipii is in the Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: jokipii@lpl.arizona.edu

1. Richardson, J. D., Kasper, J. C., Wang, C., Belcher, J. W. & Lazarus, A. J. *Nature* **454,** 63–66 (2008).
2. Decker, R. B. *et al. Nature* **454,** 67–70 (2008).
3. Stone, E. C. *et al. Nature* **454,** 71–74 (2008).
4. Burlaga, L. F. *et al. Nature* **454,** 75–77 (2008).
5. Gurnett, D. A. & Kurth, W. S. *Nature* **454,** 78–80 (2008).
6. Burlaga, L. F. *et al. Science* **309,** 2027–2029 (2005).
7. Decker, R. B. *et al. Science* **309,** 2020–2024 (2005).
8. Stone, E. C. *et al. Science* **309,** 2017–2020 (2005).
9. Jokipii, J. R., Giacalone, J. & Kota, J. *Astrophys. J.* **611,** L141–L144 (2004).
10. Lallement, R. *et al. Science* **307,** 1447–1449 (2005).
11. Ratkiewicz, R. *et al. Astron. Astrophys.* **335,** 363–369 (1998).
12. Pogorelov, N. & Matsuda, T. *J. Geophys. Res.* **10,** 237–245 (1998).
13. Opher, M., Stone, E. C. & Liewer, P. C. *Astrophys. J.* **640,** L71–L74 (2006).
14. Wang, L., Lin, R. P., Larson, D. E. & Luhmann, J. G. *Nature* **454,** 81–83 (2008).

See also pages xi and 24, and video (www.nature.com/nature/videoarchive/voyager).

ECOLOGY

Return of the niche

Mathew A. Leibold

Two ideas vie for prominence in community ecology — 'niche partitioning' and 'neutral theory'. A survey of patterns of tree abundance in tropical forest prompts fresh thinking on their respective effects.

New data and novel analyses invigorate old debates. A provocative example, published in the journal *Ecology*, comes in the form of a paper by Kelly *et al.*[1] that will spark fresh argument over the question of the factors that determine patterns of biodiversity.

Traditional explanations for the local coexistence of species hold that the balance of nature is delicately related to differences in how species interact with their local environments (their 'niches'), with populations of each species being primarily regulated by distinct environmental factors[2]. Such niche partitioning results in stable frequency dependence, in which each species increases relative to others when it is rare, and decreases when it is common[3]. This venerable view has been confronted with the contention[4,5], arising from recent modelling work, that stochastic demography and dispersal are more important, and that they allow the widespread coexistence of species with identical niches. This 'neutral theory' has provided possible explanations for the occurrence of highly diverse communities that challenge the traditional view, and has indicated ways to account for them with simple models.

The resulting clash of ideas has led to a possible synthesis that finds a place for both niche and neutrality[6,7]. The key to this synthesis is the thought that niche partitioning is increasingly less likely, and neutral dynamics more likely, as species come to resemble each other more closely. This synthesis would retain niche partitioning as a component of community structure, but would use neutral theory to resolve the matter of why niche partitioning cannot by itself explain highly diverse communities.

This is the context in which Kelly and colleagues' work[1] is set. They examined abundance patterns of trees in a highly diverse tropical forest in Mexico, and focused especially on pairwise patterns of relative abundance between closely related (congeneric) species (Fig. 1, overleaf) and less closely related species. They found that coexisting congeners had relative abundances that indicated stable frequency dependence, whereas less closely related pairs of species showed patterns consistent with neutral theory. These results are surprising — not only is it the most closely related pairs of species that show frequency dependence, rather than less closely related pairs, but also the less closely related pairs show relative abundances similar to those expected with neutral dynamics. Both of these aspects contradict the proposed synthesis described above. The authors also show that the type of frequency dependence manifested by closely related pairs of species is consistent with, and indicative of, mechanisms associated with temporally fluctuating environments. This suggests that such species-pairs coexist because one does better than another at different times, for instance

during periods of drought, rather than because they specialize on different resources.

At first glance, these results imply that closely related species show niche partitioning, whereas interactions between less closely related species are determined by neutral dynamics. However, this raises two puzzling questions.

First, how similar do species have to be to result in neutrality between them? Kelly and colleagues' findings indicate that, even in a highly diverse community, closely related congeners are not similar enough to show neutral dynamics. One possibility is that close relatedness is not reflected in ecological similarity. But this seems unlikely, because such relatedness is highly correlated with similarity of growth form, demography and general ecology.

Second, to what degree does finding relative abundance patterns consistent with neutrality really indicate that stochastic demography and dispersal regulate interactions between species-pairs? A notable point here is that Kelly *et al.* looked at pairwise relative abundance distributions, whereas most previous studies examined such distributions in entire, and often highly diverse, assemblages. Kelly and colleagues' findings tend to confirm that the fit of relative abundance distributions to neutral theory is a poor diagnostic for the importance of stochastic demography and dispersal. It may instead be that the structure of environmental variation among factors in the community itself may have random components. Some previous models of niche partitioning[8,9] have such features, and to some extent these models match the data of Kelly and colleagues.

What is surprising is that this component of niche partitioning does not also affect closely related species. The authors argue that the interactions between such species differ in type from those involving less related species because they show the signature of temporally fluctuating environments, whereas less closely related species seem less likely to be affected by this kind of niche partitioning. The mechanisms that generate temporal fluctuations in niche relations are not known, but such fluctuations may account for some aspects of highly diverse communities. However, it still seems unlikely that they can fully explain the apparent coexistence of hundreds of species in some communities, including this one.

Thus Kelly and colleagues' results provide answers to some aspects of the debate between neutralists and niche partitioners. But they leave open other issues that the tentative synthesis seemed to have resolved.

One issue is whether there is a role for neutral dynamics at all. The new findings provide surprisingly little support for such dynamics but do not rule them out. Perhaps ecological neutrality does occur among some subsets of species in these communities, but, if so, it is not closely associated with phylogenetic relatedness. The other issue is the need for an explanation for the high diversity in these

C. K. KELLY

Figure 1 | Common ground. Two adult congeneric trees — *Bursera instabilis* and *B. heteresthes* — in intimate coexistence at Chamela Biological Station in Mexico, the site studied by Kelly *et al.*[1]. Although the two species are distinct in bark coloration, they are otherwise similar in most respects.

communities. The appealing component of the synthetic view was that stochastic demography and dispersal could enhance diversity in a community with limited niches to an arbitrary degree that would depend on external factors (speciation and biotic exchanges, for example). If stochastic demography and dispersal are not important, even in a diverse community, explanations for diversity must be found elsewhere.

Kelly and colleagues' findings[1] indicate that much of the unresolved component of diversity involves the co-occurrence of comparatively unrelated species. This is a clue that the processes involved may not fall in the normal realm of conventional niche theory. This theory has focused on local dynamics, but regional factors may affect local patterns in unpredictable ways. For example, many species may be present as 'sink populations' that are maintained by dispersal from other locations where they have higher fitness as a result of different local conditions[10]. This is just one possibility, but it does suggest that regional factors may be a key to understanding highly diverse communities such as these tropical forests.

Where does all this leave us? My view is that Kelly and colleagues' paper will turn out to have a considerable impact — but as much in stimulating fresh thinking as in directly clarifying the relative roles of neutral and niche-partitioning processes. ■

Mathew A. Leibold is in the Section of Integrative Biology, University of Texas at Austin, Austin, Texas 78712, USA.
e-mail: mleibold@mail.utexas.edu

1. Kelly, C. K., Bowler, M. G., Pybus, O. & Harvey, P. H. *Ecology* **89,** 962–970 (2008).
2. Hutchinson, G. E. *Cold Spring Harb. Symp. Quant. Biol.* **22,** 415–427 (1957).
3. Chesson, P. *Annu. Rev. Ecol. Syst.* **31,** 343–366 (2000).
4. Bell, G. *Am. Nat.* **155,** 606–617 (2000).
5. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).
6. Leibold, M. A. & McPeek, M. A. *Ecology* **87,** 1399–1410 (2006).
7. Adler, P. B., HilleRisLambers, J. & Levine, J. M. *Ecol. Lett.* **10,** 95–104 (2007).
8. Nee, S., Harvey, P. H. & May, R. M. *Proc. R. Soc. B* **243,** 161–163 (1991).
9. Sugihara, G. *Am. Nat.* **116,** 770–787 (1980).
10. Mouquet, N. & Loreau, M. *Am. Nat.* **162,** 544–557 (2003).

ATMOSPHERIC CHEMISTRY

Her dark materials

Kevin Zahnle

A glitch in the history of sulphur isotopes could imply that methane emitted by the ancient biosphere created a high-altitude photochemical smog, which governed the climate in a distinctly Gaian way.

Free oxygen became abundant in Earth's atmosphere around 2.3 billion years ago, a change that was preceded or accompanied by at least three major ice ages. These ice ages themselves occurred immediately after an abrupt change in the relative abundance of isotopes of sulphur in sedimentary rocks around 2.46 billion years ago[1–3]. The sulphur isotopes indicate that the photochemistry of the atmosphere altered dramatically, and the cold climate suggests that a major greenhouse gas, probably methane, had been removed. A somewhat similar pattern in the sulphur isotope record half a billion years earlier was reported recently, also associated with an ice age, and it has been suggested that this might indicate that there had been a transient period of high atmospheric oxygen[4].

Writing in *Earth and Planetary Science Letters*, Domagal-Goldman *et al.*[5] offer a counterproposal: the sulphur isotopes and the ice age could have been caused by increased

atmospheric methane. In the authors' model, the action of ultraviolet light on methane produces a high-altitude hydrocarbon haze akin to the photochemical smog that shrouds Saturn's moon Titan. They argue that the sooty skies and methane-dependent climate were maintained by feedback with the primitive biosphere[5] — an example of Gaian management.

Sulphur is highly susceptible to reduction and oxidation, and so the history of its stable isotopes should track the history of oxygen[6]. Most chemical reactions show a slight preference for one isotope of an element over another, usually because of their differing masses. With sulphur, if a process favours ^{33}S over ^{32}S, it will favour ^{34}S twice as much. This normal sort of fractionation is called 'mass dependent'. Rarer processes that favour one isotope over another, in ways unrelated to mass, produce mass-independent fractionation (MIF)[7]. MIF is measured using the difference between the actual abundance of an isotope and the abundance that would be expected if fractionation were mass dependent. For sulphur, the third most abundant isotope is ^{33}S, and its deviation from mass-dependent fractionation is called Δ^{33}S. Since 2.3 billion years ago (that is, since the establishment of oxygen in the atmosphere), Δ^{33}S has been near zero as determined from sedimentary rocks (Fig. 1), indicating that little or no MIF has taken place. In older sediments Δ^{33}S varies considerably, and in rocks older than 2.45 billion years it is often huge[1,2].

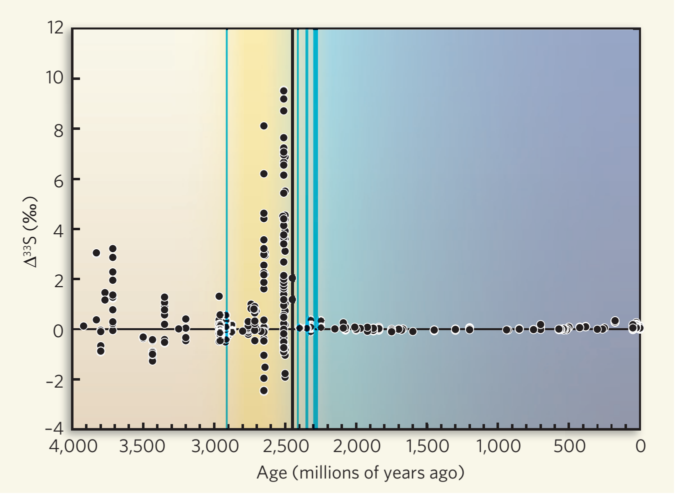


Figure 1 | Variation in ancient sulphur. **Mass-independent fractionation of ^{33}S in sedimentary rocks (Δ^{33}S measures the deviation from mass-dependent fractionation) varied wildly before 2.45 billion years ago (vertical black line) and has been near zero since, a transition that was accompanied by major ice ages (blue lines). Non-zero Δ^{33}S results from photolysis of SO_2 by ultraviolet light, and indicates the absence of atmospheric oxygen. A period of near-zero Δ^{33}S also occurred around the time of the earliest-known ice age 2.9 billion years ago[4]. Domagel-Goldman *et al.*[5] argue that increased methane created a smoggy stratosphere that blocked ultraviolet light, leading to near-zero Δ^{33}S, and blocked sunlight in general, cooling the surface of the planet. (Adapted from Fig. 1 of ref. 2.).**

The chief volcanic sulphur gas, sulphur dioxide, when exposed to ultraviolet light of wavelengths between 170 and 220 nanometres, breaks down into elemental sulphur and sulphate, with the two products having strong and opposite MIF (ref. 8). Ozone absorbs light of these wavelengths, and so in today's atmosphere only the tiny amount of sulphur dioxide that rises above the ozone layer is broken down in this way. The high levels of oxygen in our environment eventually oxidize essentially all sulphur to form sulphates, abolishing the small MIF signatures created[9].

In an atmosphere lacking ozone and other forms of oxygen, light-induced breakdown of sulphur dioxide would take place, resulting in a large MIF of sulphur. Photochemical models[9,10] predict that elemental sulphur becomes an important product when the atmosphere contains much less than one part per million oxygen and when there is a significant presence of a reduced gas such as methane. Non-zero Δ^{33}S values have been taken as an indicator of such conditions.

Recent data show a brief period around the time of the earliest-known ice age, 2.9 billion years ago, when Δ^{33}S was nearly zero[4]. One possible explanation for this is that oxygen was transiently present in the atmosphere[4]. A less extreme possibility is that both the cold climate and reduced MIF were caused by methane dropping to levels low enough to suppress the formation of elemental sulphur[10]; methane supply to the atmosphere can drop for many reasons other than the presence of oxygen. Domagal-Goldman *et al.*[5] suggest instead that high levels of methane caused both the reduced MIF and the low temperatures.

At modest levels, methane contributes to the greenhouse effect. But if there is more methane than carbon dioxide in the atmosphere, methane's interaction with ultraviolet light produces a photochemical 'soot'. As can be seen on Titan, this soot forms a brown haze high in the atmosphere that acts somewhat like ozone to reduce light-induced breakdown of sulphur dioxide and suppress MIF. Because the haze absorbs sunlight but does not insulate against heat loss, it also acts as a potent anti-greenhouse agent, which cools the surface of the planet. On Earth, methane is chiefly made by living organisms, and so for Earth's atmosphere to contain more methane than carbon dioxide requires a very active, highly methanogenic biosphere. More conventional hypotheses invoking less methane have the advantage of requiring fewer cows, metaphorically speaking.

A better understanding of the precise mechanism by which MIF of sulphur occurs in an atmosphere may allow measurements of ^{36}S (sulphur's rarest and least-studied stable isotope) to distinguish between the conflicting models[2,5]. Here, observations are well ahead of theory. Data from the 2.9-billion-year-old rocks show a different relationship between Δ^{33}S and Δ^{36}S from that seen 200 million to 400 million years later, implying that different MIF mechanisms were in operation[2]. This could be consistent with smog filtering ultraviolet light. However, sparse data from martian meteorites[11] also hint at a similarly different relationship between Δ^{33}S and Δ^{36}S, and there is no good reason to think that Mars ever had a methane-rich atmosphere.

Whether sooty or not, it is almost inevitable that a climate based on methane would be Gaian, in the loose sense of the word. Methane in the atmosphere is broken down relatively quickly, so the strengths of both its greenhouse and anti-greenhouse effects are sensitive to the rate of its production by living organisms. Domagal-Goldman *et al.*[5] point out that a global biosphere that emits more methane when warmer will self-regulate around a methane haze: increases in temperature would spur the biosphere to produce more methane, increasing the haze, and so lowering the temperature. Similarly, drops in temperature would be countered by a reduction in the haze. The opposite can also be argued: if the biosphere emits more methane when cooled, there is a stable point where the air is clear, and increasing the amount of methane would increase greenhouse warming. Gaia produces stability either way.

We are beginning to glimpse in the ancient Earth new kinds of weird planets, and habitable worlds of many colours. Life 2.5 billion years ago seems to have existed under dusty yellow skies, whereas 400 million years earlier things were stranger still: a dull brown blot where a pale blue dot should have been, ruled over by a smokestack Gaia. ■

Kevin Zahnle is at the NASA Ames Research Center, Moffett Field, California 94035, USA.
e-mail: kevin.j.zahnle@nasa.gov

1. Farquhar, J., Bao, H. & Thiemens, M. *Science* **289,** 756–758 (2000).
2. Farquhar, J. *et al. Nature* **449,** 706–709 (2007).
3. Papineau, D., Mojzsis, S. J. & Schmitt, A. K. *Earth Planet. Sci. Lett.* **255,** 188–212 (2007).
4. Ono, S., Beukes, N. J., Rumble, D. & Fogel, M. L. *South Afr. J. Geol.* **109,** 97–108 (2006).
5. Domagal-Goldman, S. D., Kasting, J. F., Johnston, D. T. & Farquhar, J. *Earth Planet. Sci. Lett.* **269,** 29–40 (2008).
6. Habicht, K. S., Gade, M., Thamdrup, B., Berg, P. & Canfield, D. E. *Science* **298,** 2372–2374 (2002).
7. Thiemens, M. H. *Annu. Rev. Earth Planet. Sci.* **34,** 217–262 (2006).
8. Farquhar, J., Savarino, J., Airieau, S. & Thiemens, M. H. *J. Geophys. Res. E* **106,** 32829–32839 (2001).
9. Pavlov, A. A. & Kasting, J. F. *Astrobiology* **2,** 27–41 (2002).
10. Zahnle, K., Claire, M. & Catling, D. *Geobiology* **4,** 271–283 (2006).
11. Farquhar, J., Kim, S.-T. & Masterson, A. *Earth Planet. Sci. Lett.* **264,** 1–8 (2007).

PHYSICAL CHEMISTRY

When molecules don't rebound

Mark Brouard

Picture a simple molecule as two balls attached together by a compressible spring. If an incoming atom strikes one end of the molecule, the spring compresses and the vibrating molecule jumps backwards. Or does it?

When investigating the dynamical behaviour of molecules and atoms, chemists often study reactions of single hydrogen atoms with hydrogen molecules, H_2. This textbook example is one of the simplest chemical processes in nature, yet in recent years a combination of high-level theory and state-of-the-art experiments has revealed a remarkable richness in the dynamics of the process[1–4]. Reporting on page 88, Greaves *et al.*[5] add a fresh twist to our understanding of this and other systems, with the identification of a new mechanism for molecular excitation in collisions of hydrogen atoms with D_2 — a hydrogen molecule in which both of the atoms are deuterium isotopes.

When a single hydrogen atom with high kinetic energy strikes a hydrogen molecule H_2 (or D_2 in Greaves and colleagues' study[5]) in its ground quantum state, two things can happen: a chemical reaction might occur, or energy might be transferred from the atom to the molecule without bonds being permanently broken or formed. In the second process, known as inelastic energy transfer, some of the kinetic energy associated with the initial translational motion of the atom and molecule is converted into vibrations and rotations in D_2. The vibrational motion in question is that associated with the rapid oscillation of the two deuterium atoms in the D_2 molecule about its centre of mass, whereas the rotational motion corresponds to end-over-end rotation of D_2 about its centre of mass.

The energy-transfer processes of molecular collisions have a central role in many areas, from atmospheric and combustion chemistry to biology, and have been studied widely. The take-home message from much of this work is that conversion of translational into vibrational (and to a smaller extent rotational) energy is an inefficient process[6] — forcing an oscillation in a molecule generally requires a lot of momentum to be transferred in a short time. Furthermore, the greater the impulse required, the more 'head-on' the collision needs to be, because such collisions maximize the possibility for momentum transfer.

In the accepted mechanism for such collisions, atoms can be thought of as solid balls, and molecules can be thought of as two balls connected by a compressible spring. When a single atom collides with a molecule, the atom that has been struck is pushed towards its molecular partner, compressing the spring-like bond. After this compression step, the atom rebounds as though it has hit a wall, whereas the molecule moves off in the direction that the incoming atom took before the collision (Fig. 1a). These collisionally excited molecules are said to be backward scattered with respect to the initially unexcited molecules. Similarly, the atom after collision is backward scattered with respect to its initial direction of travel.

But Greaves *et al.*[5] now provide convincing evidence for a different mechanism of vibrational excitation, termed a 'tug-of-war' inelastic collision. They focus on collisions of hydrogen atoms with deuterium molecules that are in their vibrational ground states. Collisional energy transfer promotes the molecules to a highly excited vibrational state, which exists in combination with one of several rotational states (including the rotational ground state). The authors find that, as the energy of the rotational state in vibrationally excited D_2 increases, the energized molecules become predominantly scattered forwards, rather than backwards. The final direction of motion of the atom is almost the same as its initial direction of travel (Fig. 1b). Contrary to received wisdom that only head-on collisions can lead to vibrational excitation, Greaves and colleagues' findings suggest that glancing collisions are also important.

The role of glancing collisions in such vibrational excitation processes had already been predicted in a theoretical study[4], but Greaves *et al.*[5] take things further by working out a possible mechanism, using classical mechanical 'trajectory' calculations. They find that production of their excited D_2 is intimately associated with the fact that the collisions are potentially reactive — that is, although they have sufficient energy to cross the energy barrier for a reaction, many of them recross the barrier to form vibrationally excited D_2 molecules instead of reaction products.

The consequences of this are dramatic. As the energetic hydrogen atoms approach D_2, a transitory bond is formed between the hydrogen and deuterium atoms, leading to a lengthening of the D_2 bond — exactly the opposite of the conventional bond-compression mechanism. Remarkably, the authors found that this D_2 stretching motion is also predicted by calculations for collisions that lead to backward scattering, which tend to be more head-on. Greaves *et al.* describe their discovery as a 'frustrated reaction' mechanism for vibrational excitation[5], because the processes of temporary bond formation and bond rupture during the collision have to be taken into account.

A crucial issue for the future will be to discover how widespread the proposed mechanism is in other chemical processes, because only a few of the collisions between hydrogen atoms and D_2 molecules lead to the vibrationally excited molecules in question[4,7]. As the authors suggest[5], it seems likely that at collision energies well above the reaction barrier,

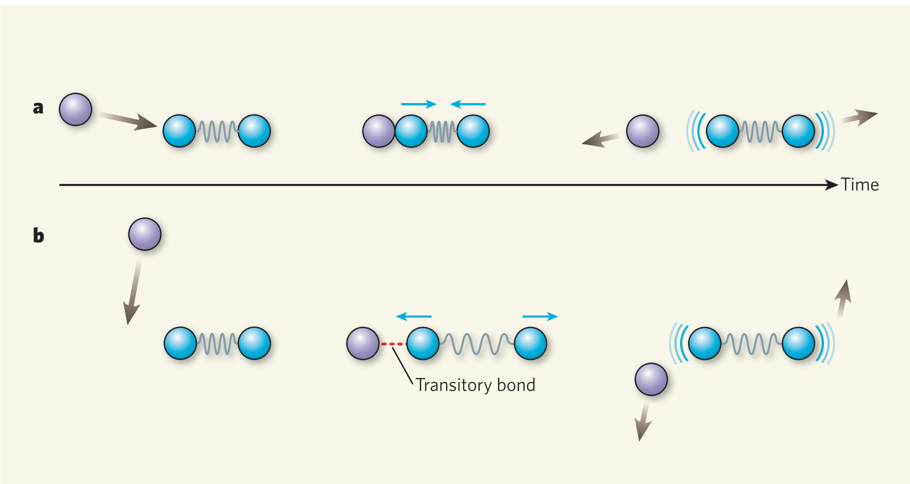


Figure 1 | Odd vibrations. When a single atom (purple) strikes a diatomic molecule (blue), some of the kinetic energy associated with the initial motion can be transferred to the molecule, causing the molecule to vibrate. **a**, In the accepted mechanism for this process, energy transfer usually occurs in a head-on collision. The colliding particles spring apart in opposite directions, with the atom moving in the opposite direction to that taken as it approached — the atom is said to be backward scattered with respect to its initial motion. **b**, Greaves *et al.*[5] propose an alternative mechanism. As the atom approaches, a transitory bond forms between it and the molecule, causing the existing molecular bond to extend. Energy transfer occurs, leading the molecule to vibrate, but the atom is forward scattered — it moves off in roughly the same direction to that taken as it approached. Brown arrows indicate movement of the single atom or the whole molecule; blue arrows indicate movement of the atoms in the molecule.

other potentially reactive systems might also undergo energy transfer in this way. In addition, there are similarities between the proposed mechanism and the 'roaming atom' mechanism identified recently[8] in the photon-induced fragmentation of formaldehyde: in both cases, high-energy molecules can undergo complex, and sometimes unpredictable, types of motion.

It will also be necessary to study the dynamics of the energy-transfer process using quantum mechanics, for comparison with the classical mechanical calculations performed by Greaves *et al.*[5]. Although it seems likely that the quantum and classical mechanisms for energy transfer will be similar, particularly for collisions that produce D_2 in higher-energy rotational states, tantalizing evidence is available suggesting that re-crossing of the reaction barrier has less of a role in quantum mechanics than in the classical calculations[4].

Against all expectations, collisions of hydrogen atoms with their molecular counterparts have succeeded in throwing up fresh surprises for eagle-eyed investigators[5]. After so many years of study, it is tempting to say that there can be little left to learn. But as Greaves and colleagues have shown, it would be foolish to bet on it. ■

Mark Brouard is in the Physical and Theoretical Chemistry Laboratory, Department of Chemistry, South Parks Road, University of Oxford, Oxford OX1 3QZ, UK.
e-mail: mark.brouard@chem.ox.ac.uk

1. Althorpe, S. C. *et al. Nature* **416,** 67–70 (2002).
2. Juanes-Marcos, J. C., Althorpe, S. C. & Wrede, E. *Science* **309,** 1227–1230 (2005).
3. Dai, D. X. *et al. Science* **300,** 1730–1734 (2003).
4. Aoiz, F. J., Bañares, L. & Herrero, V. J. *Int. Rev. Phys. Chem.* **24,** 119–190 (2005).
5. Greaves, S. G. *et al. Nature* **454,** 88–91 (2008).
6. Levine, R. D. *Molecular Reaction Dynamics* (Cambridge Univ. Press, 2005).
7. Pomerantz, A. E. *et al. J. Chem. Phys.* **121,** 6587–6590 (2004).
8. Townsend, D. *et al. Science* **306,** 1158–1161 (2004).

STEM CELLS

Tips for priming potency

Joseph F. Costello

Introducing just four specific genes into adult cells causes them to regress to an embryonic stem-cell-like state. At 1%, the efficiency of this process is low, but two tips are at hand on how to make improvements.

The American poet E. E. Cummings lionized the singular 'i', creating a persona standing apart from the crowd and, without power, scoffing at pretentious authority. Capitalist marketeers and scientists alike have repopularized the poetic 'i', perhaps innocently, in iMac and iPhone, Wii and RNAi. Biology's latest 'i' acronym is iPS cells, for induced pluripotent stem cells — antithetically one of the most potent of all cell types[1]. To produce iPS cells, drab adult cells are induced into a rejuvenated, embryonic stem (ES)-cell-like state by a puissant cocktail of just four genes — a process known as direct reprogramming. But exactly how this remarkable process works, and why it is so inefficient, has been a mystery. On page 49 of this issue, Mikkelsen *et al.*[2] explain why many of the cells get stuck in a woolly intermediate state during reprogramming, and show how to put them back on the path to pluripotency (the ability of a cell to differentiate into all adult cell types).

When thinking of cloning and reprogramming, Dolly (1997–2003), the fabled sheep begat from reprogrammed adult mammary cells[3], comes to mind. Dolly was created through nuclear transplantation, whereby the nucleus from an adult cell is transferred into an enucleated but unfertilized or only recently fertilized egg[3,4]. Both direct reprogramming and nuclear transplantation can potentially produce a viable organism from a somatic (non-germ) cell, but each technique has its advantages and disadvantages. Unlike direct reprogramming, nuclear transplantation does not use viruses, which have the potential to induce mutations, as vectors for gene transfer. But this technique requires the harvesting of precious eggs through an invasive procedure, and is not suited to efficiently deriving patient-specific ES-like cells — a goal of regenerative medicine.

The iPS cells generated by direct reprogramming, and the differentiated cells re-derived from them, not only have therapeutic potential but may also prove useful for drug discovery and for patient-specific assessment of drug toxicity. Moreover, iPS cells are 'born-again' cells that can originate from diverse adult and fetal cell types[5–8]. The humble origins of iPS cells do not incur the ethical and political controversy that plagues research into native ES cells and nuclear-transplantation-derived cells.

Although each mouse cell has some 20,000 genes, only a small subset is expressed in a specific cell type. The unique character of each cell type is established both through epigenetic mechanisms, which involve chemical modification of chromatin (complexes of DNA and histone proteins) without affecting the DNA sequence itself, and by gene transcription factors. During development, histone modification and DNA methylation package genes into 'open' or 'closed' chromatin states in a cell-type-specific way, permitting or repressing gene expression, respectively. Thus, epigenetic mechanisms dictate access of transcription factors to genes, and allow individual transcription factors to initiate distinct gene-expression programs in different cell types.

The program that establishes ES cells is of particular interest because of the pluripotency of these cells. That the overexpression of just four genes — *oct4*, *sox2*, *klf4* and c-*myc* — that encode transcription factors fully reprograms somatic cells to an ES-cell-like state, albeit with less than 1% efficiency, is quite surprising. To identify impediments to efficient reprogramming, Mikkelsen *et al.*[2] searched for crucial epigenetic modifications and gene-expression patterns in native mouse ES cells, fibroblast cells and iPS cells derived from them, and in cells trapped in intermediate states of reprogramming. They proposed that, by comparing the epigenetic and expression state of every gene in native ES cells with that in the 1% of completely reprogrammed iPS cells and that in the 99% of incompletely reprogrammed cells (intermediate cells), they could gain insight into the nature of reprogramming and the mechanisms that underlie pluripotency.

The fully reprogrammed iPS cells turn out to be remarkably similar, although not identical, to the native ES cells in terms of both gene expression and epigenetic modifications. For example, bivalent domains — chromatin that is poised for either gene expression or repression[9] — which are typical of ES cells, matched well between ES and iPS cells. Other epigenetic characteristics of ES cells observed in iPS cells include an overall larger proportion of chromatin in the open state (with the characteristic methylation of the H3 histone at its fourth lysine residue, H3K4), and localized sites of closed chromatin, characterized by H3K27 methylation, around genes specifying particular cell fates. These results are consistent with earlier reports[10–12] of similarities between ES cells and iPS cells in terms of gene expression, morphology, proliferation and ability to contribute to the germ line.

But why do the vast majority of somatic cells fail to fully reprogram? Mikkelsen and colleagues analysed three cell lines trapped in stable intermediate states, along with cells isolated at different time points during a reprogramming protocol. They identify two main impediments to reprogramming. First, stable intermediate cells and most of the cells undergoing reprogramming showed incomplete repression of genes specifying particular cell fates. Second, specific genes promoting pluripotency failed to undergo demethylation in these cells, a process that is normally necessary for these genes' expression.

Identification of these problems was important, but a more pertinent question is whether they can be fixed. The authors find

that facilitating DNA demethylation in incompletely reprogrammed cells — using a chemical inhibitor of the DNA-methylating enzyme DNMT1 — leads to a considerable response. After treatment with the inhibitor, a sizeable fraction of the stable intermediate cells exhibit three iPS-cell-like characteristics: significant demethylation of the pluripotency genes; reactivation of genes normally expressed in ES cells; and an ability to form teratomas (benign tumours) composed of all three embryonic cell layers when injected under the skin of adult mice.

There was also a time-dependent aspect to the inhibitor's effect. Incorporating this demethylating agent into the early stages of the reprogramming protocol interfered with reprogramming, whereas its addition at later stages increased the number of ES-cell-like colonies fourfold. But the demethylating agent was much more effective at enhancing reprogramming than a theoretically more specific inhibition of DNMT1 using the technique of RNA interference (RNAi). The demethylation results therefore remain open to interpretation, because the drug might have indirect or nonspecific effects, and/or the specific RNAi approach might be much less efficient at decreasing methylation.

In one of the stable intermediate cell lines, inhibitor-induced DNA demethylation alone was not sufficient to increase reprogramming efficiency. This suggests that the other potential impediment — incomplete repression of genes specifying a particular cell type — may block full reprogramming in these cells. But inhibiting the expression of several such genes with RNAi did not help. Only when RNAi-mediated inhibition of transcription factors was combined with chemically induced DNA demethylation did these intermediate cells become more amenable to reprogramming. The authors conclude that incomplete suppression of such genes and failure to demethylate DNA both interfere with reprogramming of adult cells, and that removing these impediments will enhance the efficiency of direct reprogramming. It remains to be seen whether the iPS cells generated from this more efficient protocol can contribute to the germ line when injected into a blastocyst (70–100-cell embryos), a rigorous test of their functional similarity to ES cells.

Research into direct reprogramming is advancing rapidly. Reprogramming protocols that exclude the cancer-associated gene c-*myc* have been developed[5,13]. Differentiated human cells have now been reprogrammed with the same four-gene cocktail, including cells from young and older individuals[5–7]. And the therapeutic potential of iPS cells has been demonstrated in a 'humanized' mouse, in which globin genes were replaced with human globin genes so as to model sickle-cell anaemia[14]. It will be interesting to know whether the DNA demethylating agent or inhibition of cell-type-specific factors that Mikkelsen and colleagues[2] describe will improve the efficiency of the reprogramming protocols used for human cells, and of protocols lacking c-*myc* (refs 5, 13).

iPS cells and the relatively simple methods used to generate them are of fundamental importance to biology. Reprogramming shatters the long-standing concept that the identity of differentiated adult cells is indelible. That DNA demethylation is essential for direct reprogramming is particularly interesting as this process is also strictly necessary for reprogramming by nuclear transplantation, and is a common mechanism in human cancers. Although we know very little about how DNA demethylation happens naturally, or how to manipulate it in a gene-specific way, this process clearly guides several essential cellular transition events[15]. Another puzzle is whether DNA demethylation associated with direct reprogramming involves just a few crucial genes, or occurs genome-wide. Given the current international investment in comprehensively mapping DNA methylation and other epigenetic modifications genome-wide, the 'red-hot' iPS cells will undoubtedly garner even more attention. On research into iPS cells, E. E. Cummings might have commented "into the strenuous briefness, again". ■

Joseph F. Costello is in the Department of Neurological Surgery, UCSF Helen Diller Comprehensive Cancer Center, Biomedical Sciences Program, University of California, San Francisco, San Francisco, California 94143-0875, USA.
e-mail: jcostello@cc.ucsf.edu

1. Takahashi, K. & Yamanaka, S. *Cell* **126,** 663–676 (2006).
2. Mikkelsen, T. S. *et al. Nature* **454,** 49–55 (2008).
3. Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J. & Campbell, K. H. S. *Nature* **385,** 810–813 (1997).
4. Yamanaka, S. *Phil. Trans. R. Soc. Lond. B* **363,** 2079–2087 (2008).
5. Yu, J. *et al. Science* **318,** 1917–1920 (2007).
6. Park, I.-H. *et al. Nature* **451,** 141–146 (2008).
7. Takahashi, K. *et al. Cell* **131,** 861–872 (2007).
8. Aoi, T. *et al. Science* doi:10.1126/science.1154884 (2008).
9. Bernstein, B. E. *et al. Cell* **125,** 315–326 (2006).
10. Maherali, N. *et al. Cell Stem Cell* **1,** 55–70 (2007).
11. Okita, K., Ichisaka, T. & Yamanaka, S. *Nature* **448,** 313–317 (2007).
12. Wernig, M. *et al. Nature* **448,** 318–324 (2007).
13. Nakagawa, M. *et al. Nature Biotechnol.* **26,** 101–106 (2008).
14. Hanna, J. *et al. Science* **318,** 1920–1923 (2007).
15. Hajkova, P. *et al. Nature* **452,** 877–881 (2008).

CLIMATE CHANGE

Acid test for marine biodiversity

Ulf Riebesell

Rising levels of atmospheric carbon dioxide lead to acidification of the oceans. A site in the Mediterranean, naturally carbonated by under-sea volcanoes, provides clues to the possible effects on marine ecosystems.

Much of the carbon dioxide released into Earth's atmosphere by human activities is absorbed by the oceans[1]. When dissolved in water, CO_2 forms carbonic acid. Anthropogenic carbon emissions are therefore leading to global acidification of the surface ocean[2], with uncertain consequences for marine life.

On page 96 of this issue, Hall-Spencer *et al.*[3] describe conditions off the island of Ischia near Naples, Italy (Fig. 1). Here, the release of CO_2 from under-sea volcanoes causes local acidification of sea water by as much as 1.5 pH units below the average ocean pH of 8.1–8.2. Although surrounded by a diverse rocky shore community with abundant calcareous organisms, the CO_2 venting site is impoverished in sea urchins and coralline algae, and is bare of stony corals. The shells of snails found in this area are weakened, and snail juveniles are completely absent. Are these changes a foretaste of the fate of the oceans in general?

Adverse effects of ocean acidification, particularly on organisms that build shells and skeletons from calcium carbonate, have been reported from experiments on individual species and enclosed communities[4,5]. Such experiments rely almost exclusively on abrupt and short-term changes in CO_2 concentrations, raising questions about the relevance of the observed responses to marine ecosystems exposed to high CO_2 and low pH over periods of years or decades. This includes uncertainties about the ability of marine organisms to adapt to the projected ocean acidification, and whether species sensitive to high CO_2 and low pH might be replaced by more robust forms of life without jeopardizing the overall functioning of the ecosystem.

Hall-Spencer *et al.*[3] take research in this field an important step forwards by investigating the long-term biological effects of permanent exposure to high CO_2 concentrations on a natural ecosystem. In addition to confirming laboratory-based results on individual species, they see a substantial shift in the benthic community composition, with no indication of adaptation or replacement of sensitive species by others capable of filling the same ecological niche. As predicted from previous work[6], however, there are winners as well as losers in ocean acidification and carbonation. Although calcareous groups generally decline in abundance or vanish completely, photosynthetic groups such as sea grasses and brown algae benefit from the higher CO_2 availability by increasing their biomass.

R. RODOLFO-METALPA

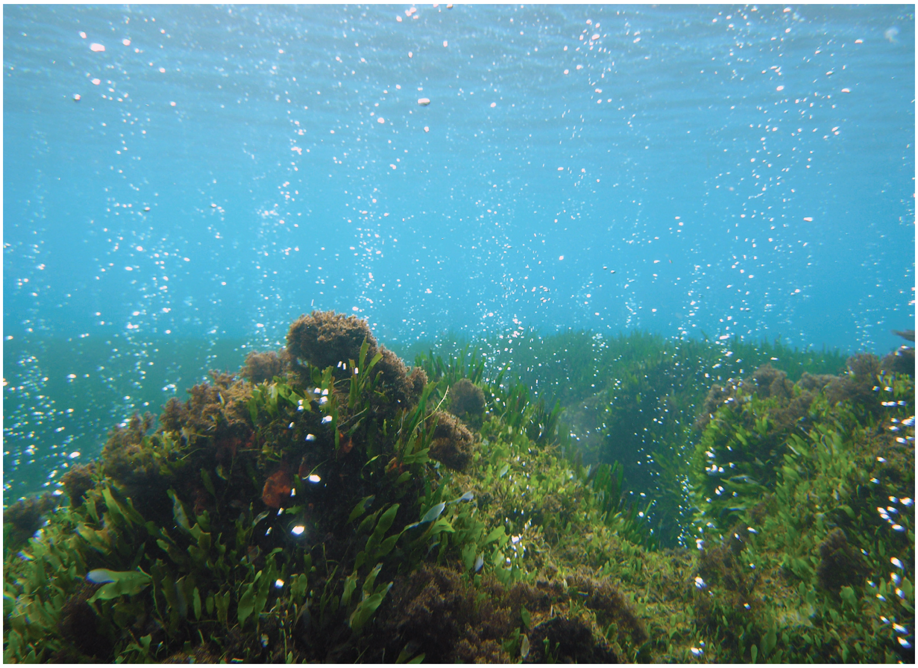


Figure 1 | Sparkling sea water. **Venting of volcanic CO_2 at a Mediterranean site off the island of Ischia provides the opportunity to observe changes in the community structure of a rocky shore ecosystem along gradients of decreasing pH close to the vents. Groups such as sea urchins, coralline algae and stony corals decline in abundance or vanish completely with decreasing pH. Sea grasses and brown algae benefit from elevated CO_2 availability close to the vent by increasing their biomass. Similar high CO_2/low pH conditions are on the verge of progressively developing ocean-wide through the uptake of fossil-fuel CO_2 by the surface ocean.**

This study[3] is a compelling demonstration of the usefulness of natural CO_2 venting sites in assessing the long-term effects of ocean acidification on sea-floor ecosystems, an approach that undoubtedly needs to be further explored. But there are considerable differences between such systems and the situation arising from global-scale ocean acidification caused by rising atmospheric CO_2. For example, temporal and spatial variability in CO_2 and pH perturbations, induced in part by changes in the direction and intensity of water currents, complicate the determination of a reliable dose–response relationship. Large but short-term variation in pH may itself be stressful to some organisms owing to the extra physiological burden of acclimating to ever-shifting conditions. In addition, mobile species and planktonic stages continually move or are carried into the venting area, providing a supply of organisms previously unexposed to high CO_2 and low pH. This further complicates the extrapolation of CO_2 effects from volcanic vents to global-scale ocean acidification. Invasion of non-adapted organisms may also cause short-term stress to those organisms, possibly amplifying the range of high-CO_2 responses.

In the case of unabated CO_2 emissions, ocean acidification may develop to pose an unprecedented threat to marine life. Our understanding of the processes that underlie its observed effects on ecosystems and biogeochemistry is still rudimentary, as is our ability to forecast its impacts. There is an urgent need to develop tools to assess and quantify such impacts across the entire range of biological responses, from subcellular regulation to ecosystem reorganization, and from short-term physiological acclimation to evolutionary adaptation.

Hall-Spencer *et al.*[3] provide independent support for conclusions, reached by experimental studies, that ocean acidification can cause a loss of biodiversity and trigger shifts in ecosystem structure and function[4–6]. They also demonstrate that, although natural CO_2 venting sites are not precise analogues of global-scale ocean acidification, they can provide essential information about high-CO_2 effects on spatial and temporal scales, which are otherwise difficult to address. Tackling this emerging threat to marine biota calls for a coordinated research effort and requires "a coherent global vision ... to better determine the impacts of climate change on marine systems"[7]. ■

Ulf Riebesell is at the Leibniz Institute of Marine Sciences (IFM-GEOMAR), 24105 Kiel, Germany.
e-mail: uriebesell@ifm-geomar.de

1. Sabine, C. L. *et al. Science* **305,** 367–371 (2004).
2. Orr, J. C. *et al. Nature* **437,** 681–686 (2005).
3. Hall-Spencer, J. M. *et al. Nature* **454,** 96–99 (2008).
4. Fabry, V. J., Seibel, B. A., Feely, R. A. & Orr, J. C. *ICES J. Mar. Sci.* **65,** 414–432 (2008).
5. Kleypas, J. A. *et al. Impacts of Ocean Acidification on Coral Reefs and Other Marine Calcifiers* workshop rep. 18–20 April 2005, St Petersburg, Florida, www.ucar.edu/communications/Final_acidification.pdf (2006).
6. Royal Society *Ocean Acidification Due to Increasing Atmospheric Carbon Dioxide* Policy Document 12/05 (Royal Society, London, 2005).
7. Richardson, A. J. & Poloczanska, E. S. *Science* **320,** 1294–1295 (2008).

See also page 16.

50 & 100 YEARS AGO

50 YEARS AGO

At a meeting of the Linnean Society on July 1, attended by members of the Darwin and Wallace families, representatives of other societies and institutions and members of the Linnean Society, the president, Dr. C. F. A. Pantin, unveiled a plaque in the meeting room commemorating the centenary of the reading before the Society on July 1, 1858, of the joint communication by Charles Darwin and Alfred Wallace on their theory of evolution by natural selection. At the meeting a hundred years ago neither Darwin nor Wallace was present: Darwin because of family bereavement and illness, and Wallace was still in Ternate. The papers were communicated by Sir Charles Lyell and Dr. (later Sir) J. D. Hooker ... Hooker, writing to Francis Darwin at a later date giving an account of the meeting, said " ... The interest excited was intense, but the subject too novel and too ominous for the old School to enter the lists before armouring. It was talked over after the meeting, 'with bated breath' ..."
From *Nature* 5 July 1958.

100 YEARS AGO

The list of honours issued on the occasion of His Majesty's birthday includes the name of a few men distinguished for their work in pure or applied science ... Some reference has been made in the daily papers to the ratio of honours awarded to naval and military men, the suggestion being that the Army receives an undue share of these distinctions. With the demands of the two services for recognition we are not concerned, but the question induces us to ask what ratio exists between the award of honours to men who devote their lives to work which promotes the scientific progress of the country and those who do not? ... Probably the reason is that ministers and officials who are chiefly concerned with the affairs of State and Court live in a world in which science and the results of science are almost unknown.
From *Nature* 2 July 1908.

Vol 454 | 3 July 2008 | doi:10.1038/nature07056

ARTICLES

Dissecting direct reprogramming through integrative genomic analysis

Tarjei S. Mikkelsen[1,2], Jacob Hanna[4], Xiaolan Zhang[1], Manching Ku[5], Marius Wernig[4], Patrick Schorderet[4], Bradley E. Bernstein[1,5,6], Rudolf Jaenisch[3,4], Eric S. Lander[1,3,4,7] & Alexander Meissner[1,8]

Somatic cells can be reprogrammed to a pluripotent state through the ectopic expression of defined transcription factors. Understanding the mechanism and kinetics of this transformation may shed light on the nature of developmental potency and suggest strategies with improved efficiency or safety. Here we report an integrative genomic analysis of reprogramming of mouse fibroblasts and B lymphocytes. Lineage-committed cells show a complex response to the ectopic expression involving induction of genes downstream of individual reprogramming factors. Fully reprogrammed cells show gene expression and epigenetic states that are highly similar to embryonic stem cells. In contrast, stable partially reprogrammed cell lines show reactivation of a distinctive subset of stem-cell-related genes, incomplete repression of lineage-specifying transcription factors, and DNA hypermethylation at pluripotency-related loci. These observations suggest that some cells may become trapped in partially reprogrammed states owing to incomplete repression of transcription factors, and that DNA de-methylation is an inefficient step in the transition to pluripotency. We demonstrate that RNA inhibition of transcription factors can facilitate reprogramming, and that treatment with DNA methyltransferase inhibitors can improve the overall efficiency of the reprogramming process.

Mouse and human cells can be reprogrammed to pluripotency through ectopic expression of defined transcription factors[1–9] ('direct reprogramming'). Generation of such induced pluripotent stem (iPS) cells may provide an attractive source of patient-specific stem cells (reviewed in refs 10, 11). However, the mechanism and nature of molecular changes underlying the process of direct reprogramming remain largely mysterious[11]. It is a slow and inefficient process that currently requires weeks, with most cells failing to reprogramme[2,9,12–14]. A clearer understanding of the process would enable development of safer and more efficient reprogramming strategies, and might shed light on fundamental questions concerning the establishment of cellular identity.

To identify possible obstacles to reprogramming and to use this knowledge to devise ways to accelerate the transition to full pluripotency, we undertook a comprehensive genomic characterization of cells at various stages of the reprogramming process. The characterization involved gene expression profiling, chromatin state maps of key activating and repressive marks (histone H3 K4me3 and K27me3) and DNA methylation analysis.

Response to reprogramming factors

We first studied the response of lineage-committed cells to ectopic expression of the four reprogramming factors Oct4 (also known as Pou5f1), Sox2, Klf4 and c-Myc. Because most induced cells fail to achieve successful reprogramming, we reasoned that genomic characterization might yield insights into the basis of the low overall efficiency of the method.

To eliminate heterogeneity caused by differential viral integration, we studied mouse embryonic fibroblasts (MEFs) isolated from chimaeric mice that had been generated from an iPS cell line carrying integrated doxycycline (Dox)-inducible lentiviral vectors with the four reprogramming factors and a *Nanog–GFP* (green fluorescent protein) reporter gene[13,15]. We induced the expression of the reprogramming factors and obtained gene expression profiles at days 4, 8, 12 and 16 (Supplementary Data). Fluorescence-activated cell sorting (FACS) analysis on day 16 showed that ~20% of the cells stained positive for the stem-cell marker SSEA1, but only ~1.2% had achieved complete reprogramming, as indicated by activation of the Nanog–GFP reporter (Supplementary Fig. 1) and consistent with previous reports[13,14].

The immediate response to induction of the reprogramming factors (>3-fold change by day 4) is characterized by de-differentiation from the wild-type MEF state and upregulation of proliferative genes. De-differentiation is evident in a significant decrease (5–40-fold) in expression levels of typical mesenchymal genes expressed in MEFs (for example, *Snai1* and *Snai2*). The proliferative response is evident in upregulation of genes with functions such as DNA replication (*Poli*, *Rfc4* and *Mcm5*) and cell cycle progression (*Ccnd1* and *Ccnd2*); this response may be consistent with expression of reprogramming factor c-Myc[10,16].

We also detected a strong increase in the expression of stress-induced and anti-proliferative genes. In particular, we detected a sustained 5–10-fold upregulation of *Cdkn1a* and *Cdkn2a*, which encode cyclin-dependent kinase (CDK) inhibitors that are key effectors of multiple differentiation and tumour suppressor pathways. *Cdkn1a* is a downstream target of the reprogramming factor Klf4 (ref. 17), whereas *Cdkn2a* is known to be activated by deregulated c-Myc expression[18]. This response was followed by gradual upregulation of genes associated with differentiating MEFs (*Pparg*, *Fabp4* and *Mgp*) on days 12–16. This suggests that induction of the reprogramming factors

[1]Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Division of Health Sciences and Technology, [3]Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [4]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. [5]Molecular Pathology Unit and Center for Cancer Research, Massachusetts General Hospital, Charlestown, Massachusetts 02129, USA. [6]Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA. [7]Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02114, USA. [8]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA.

triggers normal 'fail-safe' mechanisms that act to prevent uncontrolled proliferation, which may prevent the majority of cells from reaching a stably de-differentiated state.

We also detected strong upregulation of lineage-specific genes from unrelated lineages. These include axon guidance factors (*Epha7* and *Ngef*), epidermal proteins (*Krt14*, *Krt16*, *Ivl* and *Sprr1a*) and glomerular proteins (*Podxl*). We speculate that this gene activation reflects responses to the reprogramming factors Sox2 and Klf4, which, independent of their roles in embryonic stem cell regulation, function in neural, epidermal and kidney differentiation[10,17].

Pluripotent cell lines

We next studied the changes to gene expression patterns and epigenetic states seen in successfully reprogrammed iPS cells. We analysed three cell lines: MEF-derived iPS cells carrying an Oct4–GFP reporter (MCV8.1; corresponding to subclone 8.1 in ref. 12); mature-B-lymphocyte-derived iPS cells carrying a Nanog–GFP reporter (B-iPS)[15]; and wild-type embryonic stem cells (V6.5)[19].

We found that the genome-wide expression profiles of Oct4- or Nanog-iPS cells derived from different cell types and systems are highly similar, but not identical, to wild-type embryonic stem cells (Fig. 1), consistent with recent studies of independent cell lines[2,4,9,20]. For example, the iPS and embryonic stem cell lines share high expression levels of genes related to maintenance of pluripotency and self-renewal such as *Oct4*, *Sox2*, *Nanog*, *Lin28*, *Zic3*, *Fgf4*, *Tdgf1* and *Rex1* (also known as *Zfp42*), and low expression levels for most lineage-specifying transcription factors and other developmental genes. Consistent with the characteristically short cell cycle of embryonic stem cells, the iPS cells show low expression of cyclin D (*Ccnd1* and *Ccnd2*)[21].

To determine whether iPS cells have also regained embryonic-stem-cell-like chromatin states, we generated genome-wide maps showing the location of H3K4me3 and H3K27me3 from the MEF-derived MCV8.1 cell line using ChIP-Seq. Previously we described the differences in these chromatin modifications between wild-type embryonic stem cells and MEFs[22]. In embryonic stem cells, virtually all high-CpG promoters (HCPs) are enriched with H3K4me3; a subset of these HCPs, associated with repressed developmental genes, are also enriched with H3K27me3 ('bivalent'). In MEFs, most HCPs that are bivalent in embryonic stem cells resolve to become monovalent (H3K4me3- or H3K27me3-only). Some pluripotency- and germline-specific genes show loss of both H3K4me3 and H3K27me3 in somatic cells, and this correlates with DNA hypermethylation (ref. 23, and A.M. *et al.*, unpublished observations).

The chromatin state maps of the iPS cell line MCV8.1 are markedly similar to those of embryonic stem cells both near promoters and in intergenic regions (Fig. 2 and Supplementary Figs 2–6). Most (>97%) HCPs that lack H3K4me3-enrichment in MEFs have regained this mark in MCV8.1 cells. At all pluripotency- and germline-specific genes examined, the promoters have regained H3K4me3-enrichment and show DNA hypomethylation (Fig. 3). At genes encoding lineage-specific transcription factors that are bivalent and transcriptionally silent in embryonic stem cells, the bivalent pattern is typically re-established (~80% of HCPs classified as bivalent in wild-type embryonic stem cells, and ~95% of loci encoding key developmental transcription factors; Fig. 2b–d, g).

We conclude that direct reprogramming to a pluripotent state involves re-activation of endogeneous pluripotency-related genes, establishment of an 'open' chromatin state (as indicated by genome-wide H3K4me3 enrichment and DNA de-methylation), and comprehensive Polycomb-mediated repression of lineage-specifying genes (as indicated by bivalent chromatin states involving H3K27me3-enrichment).

Partially reprogrammed cell lines

Only a subset of the stably de-differentiated cells obtained in the absence of drug selection show evidence of complete reprogramming to a pluripotent state. Previously we derived clonal cell lines that can be maintained in relatively stable 'partially reprogrammed' states in the absence of drug selection[12]. We reasoned that characterizing such cells might help to identify key barriers in the late stages of the process. Accordingly, we studied three partially reprogrammed independent cell lines established during attempts to reprogramme MEFs or mature B lymphocytes (Figs 1–3).

MCV8. This cell line, which corresponds to subclone 8 from ref. 12, was established during our attempt to reprogramme MEFs carrying an Oct4–GFP reporter with constitutive retroviruses. It produces heterogeneous cultures of cells with mainly fibroblast-like morphology, with ~20–30% positive for the stem cell marker SSEA1 (Supplementary Figs 7 and 8) and occasional interspersed embryonic-stem-cell-like colonies at late passages. Multiple secondary subclones from these embryonic-stem-cell-like colonies have been shown to establish homogeneous GFP-positive iPS cell lines

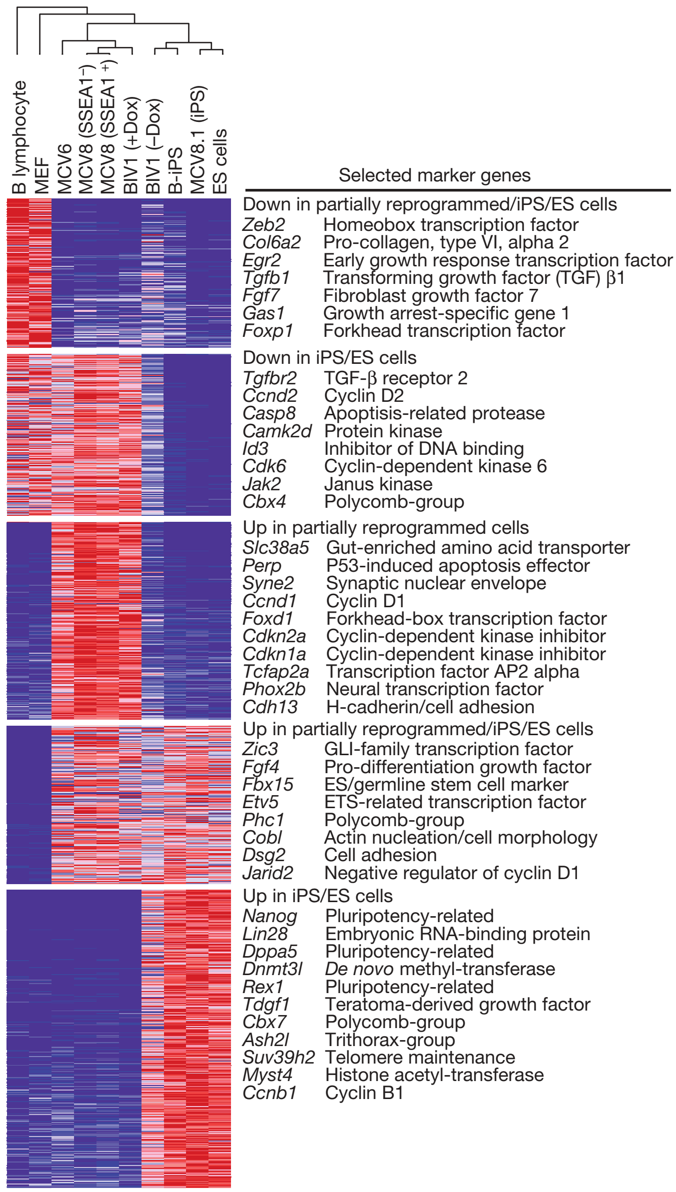


Figure 1 | Gene expression profiling. Relative expression levels across differentiated, partially reprogrammed and pluripotent cell populations. The dendrogram was generated by complete linkage hierarchical clustering using Pearson correlation on all measured genes. Only genes with at least twofold difference between any pair of samples from different classes are shown in the heat map. Red, white and blue indicate higher, identical and lower relative expression, respectively. ES cells, embryonic stem cells.

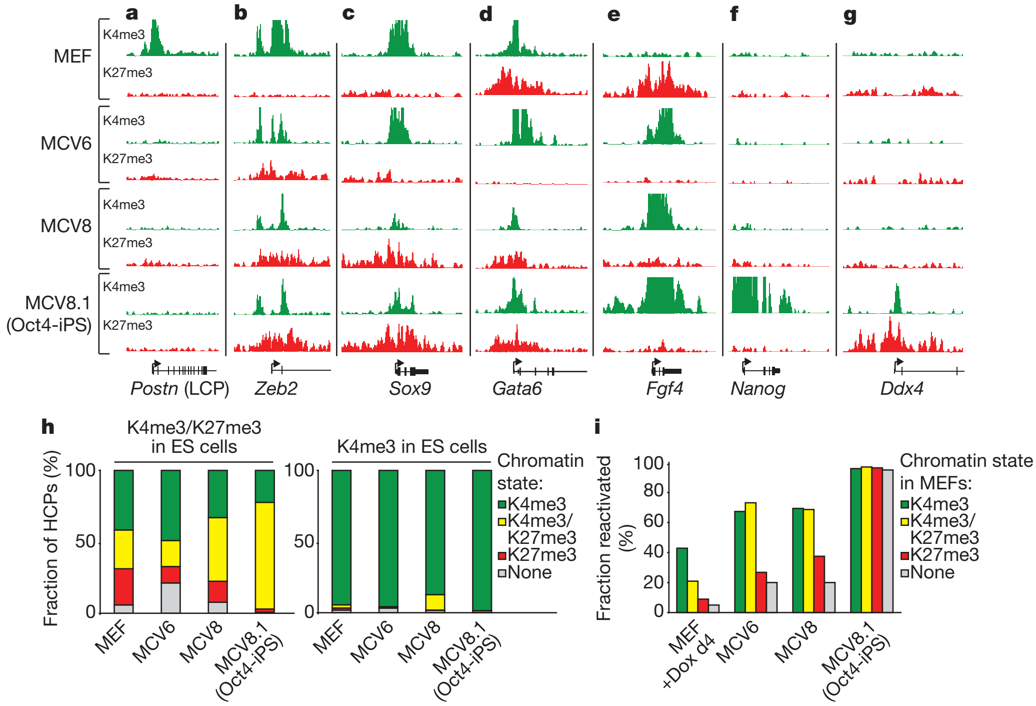


Figure 2 | Chromatin state maps. **a**, Loss of H3K4me3 correlates with inactivation of MEF-specific low-CpG promoters (LCPs), such as that of *Postn* (periostin), during reprogramming. **b**, The transcription factor *Zeb2* is marked by H3K4me3 and expressed in MEFs, but gains H3K27me3 and is silenced in partially and fully reprogrammed cells. **c**, The mesoderm/neural-crest transcription factor *Sox9* is marked by H3K4me3 only and remains active in MCV6. **d**, The endodermal transcription factor *Gata6* inappropriately lost H3K27me3 and is activated in MCV6 cells. **e**, The autocrine growth factor *Fgf4* loses H3K27me3, gains H3K4me3 and becomes highly expressed in both partially and fully reprogrammed cells. **f**, The pluripotency gene *Nanog* gains H3K4me3 and is active only in iPS cells. **g**, The germline-specific gene *Ddx4* gains H3K4me3 and H3K27me3 in iPS cells only, and remains poised for activation in germ cells. **h**, Chromatin states for high-CpG promoters (HCPs) in MEFs and reprogrammed cells, conditional on their state in embryonic stem cells. **i**, Fraction of genes with HCPs expressed in embryonic stem cells, but not wild-type MEFs, that have been re-activated in cells at various stages of reprogramming, conditional on their chromatin state in MEFs. Most HCPs marked by H3K27me3 only or by neither mark are not re-actived in partially reprogrammed cells. d4, day 4.

(including the MCV8.1 line characterized above[12]). Proviral integration patterns showed that the same parental cells in the MCV8 population gave rise to both GFP-positive and -negative cells, suggesting that complete reprogramming depends on stochastic epigenetic events[11,12].

The gene expression patterns of MCV8 cells are clearly distinct from both MEFs and iPS cells (Fig. 1). MCV8 cultures show down-regulation of both structural genes (*Col1a1* and *Col1a2*) and regulatory factors (*Snai1*, *Snai2* and *Zeb2*) expressed in MEFs, upregulation of some lineage-specific genes with neural, epidermal or endodermal functions (presumably as a consequence of Sox2 and Klf4 expression), and particularly high expression of proliferative genes. Interestingly, high levels of expression can also be detected for several of the CDK inhibitors (*Cdkn1a* and *Cdkn2a*) induced by the reprogramming factors. It is unclear how the partially reprogrammed cells have escaped the presumed anti-proliferative effects of these genes, but possible explanations include compensation by overexpression of proliferative genes, repression of differentiation pathways (MCV8 is cultured in the presence of the differentiation inhibitor LIF and expresses the LIF receptor at 2–3-fold higher levels than embryonic stem cells) or transformation (but we note that MCV8 cells have not lost the ability to re-differentiate, see below).

The pattern of re-activation of genes expressed in embryonic stem cells in MCV8 is strongly correlated with chromatin state in MEFs (Fig. 2i). Several genes related to self-renewal and proliferation of embryonic and adult stem cells show re-activation, including the autocrine growth factor *Fgf4* (ref. 24) and the transcription factor *Zic3* (ref. 25), but genes directly related to pluripotency show low or undetectable expression. Of HCPs that are enriched with H3K4me3 in MEFs but are not expressed at detectable levels, most (~70%) are

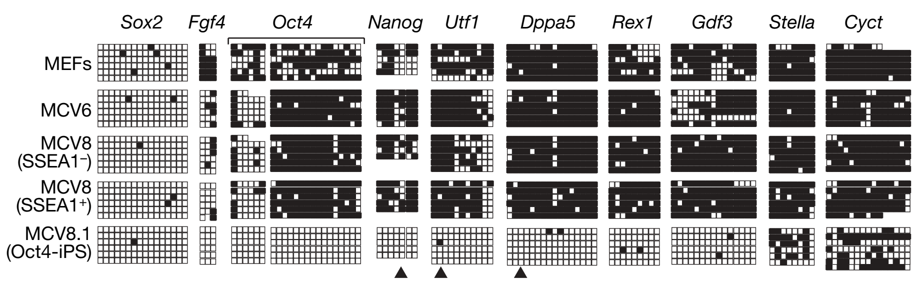


Figure 3 | DNA methylation analysis. Bisulphite sequencing of promoters or enhancers with Oct4/Sox2 binding sites near pluripotency-related and germ-cell-specific (*Stella* and *Cyct*) genes, as catalogued in ref. 23. Empty squares indicate unmethylated and filled squares methylated CpG dinucleotides. Most assayed sites are hypermethylated in differentiated and partially reprogrammed cells. *Sox2* is enriched with H3K27me3 in non-pluripotent cells and accordingly hypomethylated in all cell types. Triangles show sites used for COBRA analysis (see text).

re-activated in MCV8. In contrast, transcriptionally silent HCPs that are enriched in MEFs for H3K27me3 only or for neither mark are significantly less likely to be re-activated (~35% and ~20%, respectively; $P_{\mathrm{Fisher}} < 10^{-6}$).

There are notable differences in the chromatin states of MCV8, MEFs and MCV8.1 iPS cells (Fig. 2). Examining HCPs that are bivalent in embryonic stem cells demonstrates that MCV8 cells show bivalent chromatin structures at 70% more of these loci ($n = 1,467$) than seen in the MEFs ($n = 859$), but at ~40% fewer than in MCV8.1 iPS cells ($n = 2,360$); this is consistent with partial de-differentiation (~88% of the bivalent loci in MCV8 are also bivalent in MCV8.1). There are many more HCPs that lack H3K4me3 and H3K27me3 in MCV8 than in MCV8.1 ($n = 311$ versus 31), and these genes include the majority of pluripotency- and germ-cell-specific loci. Using bisulphite sequencing, we confirmed that this chromatin state correlates with DNA hypermethylation (Fig. 3).

We initially sorted MCV8 cells into SSEA1-positive and -negative cells and analysed them separately. However, we found no major differences in expression levels or DNA methylation patterns between the two fractions (Figs 1 and 3; Supplementary Data). Moreover, when the two subpopulations were cultured separately, both reverted to a heterogeneous state within 1–2 passages (Supplementary Fig. 9). Similar results were obtained from sorting by major histocompatibility complex surface expression, which decreases on reprogramming (Supplementary Fig. 10; Supplementary Data). Thus, although these surface markers may provide some enrichment for cells that are amenable to full reprogramming[14], they do not seem to discriminate between significantly different cell states within MCV8 cultures.

MCV6. This cell line was also established during our attempt to reprogramme Oct4–GFP MEFs (subclone 6 from ref. 12). It produces homogeneous cultures with compact colonies and embryonic-stem-cell-like morphology (Supplementary Fig. 8). It differs from MCV8 in that it has different proviral integrations and has never spontaneously given rise to fully reprogrammed cells (Supplementary Fig. 7).

The gene expression profile and chromatin state maps from MCV6 are largely similar to those of MCV8, but we found two notable differences. First, MCV6 has fewer genes with bivalent chromatin signatures, and a disproportionately large fraction of HCPs with neither H3K4me3- nor H3K27me3-enrichment (7% versus ~2.5% in MEFs and MCV8). Second, MCV6 expresses high levels of several lineage-specifying transcription factors that are expressed at low or undetectable levels in MCV8 or iPS cells, including *Sox9* (Fig. 2c) and *Gata6* (Fig. 2d). The latter observation suggests that MCV6 may have become trapped in a more differentiated state than MCV8.

BIV1. This cell line was established during our attempt to reprogramme B lymphocytes with inducible lentiviral vectors[15]. It had lost surface expression of all common lymphoid markers and did not require any lymphoid cytokines for growth, but also showed no evidence of achieving complete reprogramming during 50 days of continuous Dox-mediated viral expression (as judged by the absence of SSEA1- or GFP-positive cells). After Dox withdrawal and loss of any detectable viral expression (see below), the cells continued to proliferate with a more fibroblast-like morphology and, after more than ten additional days in culture, spontaneously gave rise to some GFP-positive embryonic-stem-cell-like colonies, but at a lower frequency than MCV8 (Supplementary Figs 8 and 11).

The gene expression profile and chromatin state maps from BIV1 cells grown with Dox show notable similarities to those of MCV8, including: downregulation of lineage-specific genes, such as the B lymphocyte master regulator *Pax5*; high expression of proliferative genes; activation of neural and epidermal genes; low levels of H3K4me3 and H3K27me3 enrichment relative to embryonic stem cells, consistent with DNA hypermethylation (see below); and incomplete activation of pluripotency-related loci (Fig. 1; Supplementary Figs 2–6 and 12). Notably, the expression profiles of BIV1, MCV8 and MCV6 are more similar to each other ($r^2 > 0.9$ for any pair) than to the lineage-committed cell types from which they originated or to any of the pluripotent cell types ($r^2 < 0.8$ for any pair; Fig. 1). This suggests that the three cell lines may represent relatively common intermediate states induced by the four reprogramming factors (Oct4, Sox2, Klf4 and c-Myc). (The three lines also show expression of *Fbx15*, suggesting that they may be similar to the *Fbx15*-selected cells obtained during initial attempts to generate iPS cells[7].)

By comparing the expression profiles of BIV1 cultures before and after Dox withdrawal, we found that Dox withdrawal resulted in: upregulation of mesenchymal extracellular matrix genes (*Col1a1* and *Col2a1*), consistent with the shift to a more fibroblast-like morphology; downregulation of most inappropriately expressed neural and epidermal genes, which is consistent with these genes being induced by overexpression of Sox2 or Klf4; and upregulation of some iPS and embryonic-stem-cell-specific genes (*Dppa5* (also known as *Dppa5a*), *Lin28* and *Dnmt3l*), which is consistent with the eventual emergence of rare GFP-positive colonies. Thus, continuous overexpression of the reprogramming factors may paradoxically have stabilized BIV1 cells in its partially reprogrammed state.

In summary, the three partially reprogrammed cell lines appear to represent similar (but distinct) cell states that emerge at an intermediate stage in the direct reprogramming process. The states are characterized by: re-activation of genes related to stem cell renewal and maintenance, but not pluripotency; incomplete repression of lineage-specific transcription factors; and incomplete epigenetic remodelling, including persistent DNA hypermethylation.

Inhibition of Dnmt1 accelerates reprogramming

Because the partially reprogrammed cell lines show DNA hypermethylation at pluripotency-related genes, we hypothesized that loss of DNA methylation (or a closely linked epigenetic mark, such as H3K9 methylation[26]) is a critical and inefficient step in the transition from a partially reprogrammed state to pluripotency.

Partially reprogrammed cell lines. We tested this notion by treating cells with the DNA methyltransferase inhibitor 5-aza-cytidine (AZA) and found that it induced a rapid and stable transition to a fully reprogrammed iPS state. We initially studied SSEA1-positive MCV8 cells, treating them with AZA for 48 h and monitoring the subsequent appearance of GFP-positive cells (Fig. 4a and Supplementary Fig. 7). GFP-positive cells appeared at a frequency of 7.5% after one passage, compared to 0.25% in untreated cells. After five passages, GFP-positive cells comprised 77.8% of the treated population, whereas the proportion in untreated cells remained stably low (0.41%). We obtained similar results when treating the SSEA1-negative fraction. (When untreated cells from the fifth passage were subsequently treated with AZA, GFP-positive cells appeared at a similar rate as in the initial treatment; Fig. 4b.) We also found robust induction of the GFP reporter after AZA treatment of BIV1 (−Dox) cells (Fig. 4a and Supplementary Fig. 13a).

We evaluated the cellular state and developmental potency of the GFP-positive MCV8 and BIV1 cells obtained after AZA treatment and FACS. Both populations stained positive for the stem-cell marker SSEA1. Combined bisulphite restriction analysis (COBRA) revealed significant de-methylation of CpGs near the pluripotency-related genes *Dppa5*, *Nanog* and *Utf1* (Supplementary Fig. 14), implying that re-activation was not limited to the GFP-tagged reporters. The viral transgenes showed low or undetectable expression levels (Fig. 4c, d) indicating that AZA treatment did not interfere with viral silencing, which is required for full reprogramming[9], and that the emergence of GFP-positive cells was not caused by viral re-activation. Finally, subcutaneous injection into severe combined immunodeficiency (SCID) mice led to teratoma formation in 3–4 weeks (Fig. 4e), demonstrating that the GFP-positive cells had undergone a stable transition to the pluripotent state. (Untreated MCV8 or BIV1 cells did not generate teratomas in the same time frame.)

To exclude nonspecific effects of AZA, we treated MCV8 cells with small interfering RNAs (siRNAs) or lentiviral short hairpin RNAs (shRNAs) against *Dnmt1*, which also led to the appearance GFP-positive cells within one passage (up to 1.7%; Supplementary Fig. 13b–d). We conclude that transient inhibition of Dnmt1 is sufficient to transition MCV8 and BIV1 cells rapidly from a partially reprogrammed state to a pluripotent state.

Populations of lineage-committed cells. We next used the chimaera-derived Nanog–GFP MEFs (described previously) to test whether AZA treatment could increase the overall reprogramming efficiency. The cells were grown in the presence of Dox from day 1, and AZA was administered for 48 h starting on day 4, 6 or 8. The reprogramming efficiency was determined by counting embryonic-stem-cell-like colonies at day 14 (Fig. 4f, g).

We found that starting AZA treatment on days 4 and 6 led to high cell death and no overall gain in efficiency. The cell death may reflect the fact that most cells are still in a differentiated state: genome-wide hypomethylation is known to induce apoptosis in differentiated cells,

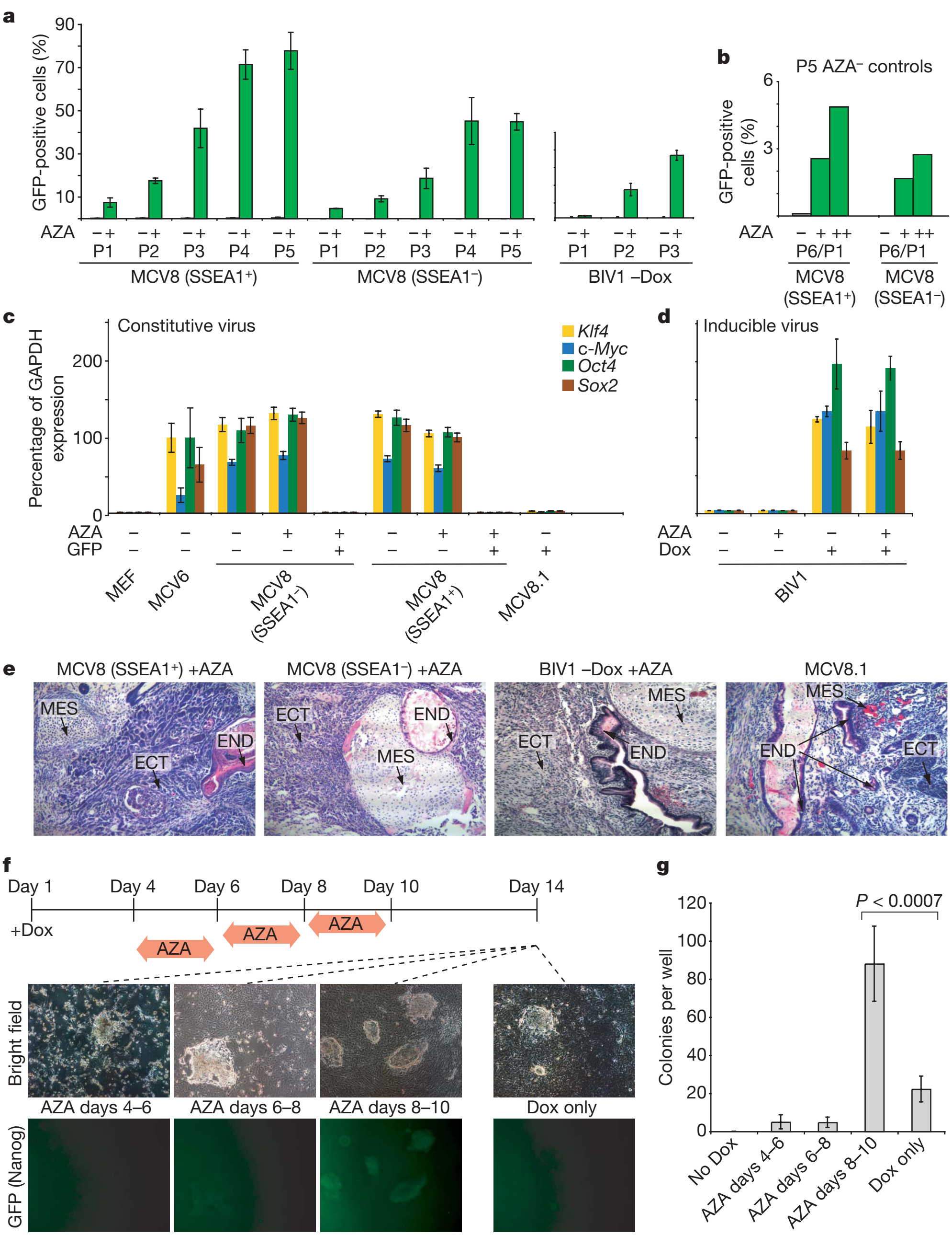


Figure 4 | Inhibition of Dnmt1 accelerates the transition to pluripotency. **a**, MCV8 (sorted by FACS using a SSEA1-specific antibody) and BIV1 (−Dox) were either exposed to AZA for 48 h (green) or kept in regular embryonic stem cell medium (grey). The number of Oct4–GFP-positive cells was analysed over multiple passages (P) by FACS. **b**, Untreated MCV8 control cells from passage 5 were subsequently subjected to AZA treatment for 48 h (+) or 120 h (++), and resulting Oct4–GFP-positive cells were counted after one passage. P6/P1, total passage 6/passage 1 after AZA treatment. **c**, AZA treatment does not influence retroviral expression levels. **d**, AZA treatment has no influence on lentiviral expression in uninduced or induced BIV1 cells. **e**, Pluripotency of all AZA-treated lines and MCV8.1 was demonstrated by teratoma formation. ECT, ectoderm; END, endoderm; MES, mesoderm. **f**, Overall efficiency of AZA treatment. Nanog–GFP MEFs were plated on 6-well plates (4 wells per time point with Dox, and 2 wells without). Cells were treated with AZA during one of the indicated intervals. On day 14, colony formation was analysed by fluorescence microscopy (representative panels are shown). **g**, Number of alkaline-phosphatase-positive, embryonic-stem-cell-like colonies obtained from each treatment. AZA treatment during days 8–10 resulted in a ~4-fold increase in efficiency over untreated controls. For **a**, **c**, **d** and **g**, error bars show standard deviations ($n = 2$, 2, 2 and 4, respectively).

whereas embryonic stem cells are resistant[27–29]. In contrast, there was a consistent fourfold increase in the number of embryonic-stem-cell-like colonies in the cultures treated with AZA starting on day 8 ($P < 0.007$; *t*-test). Moreover, most (>95%) embryonic-stem-cell-like colonies were GFP-positive in the treated cells, whereas only a minority (<25%) were GFP-positive in the untreated controls (a proportion consistent with refs 9, 12–14). Whereas early AZA treatment is counter-productive to reprogramming, there may be a sufficient number of partially reprogrammed cells in the population to outweigh its cytotoxic effect.

We conclude that de-methylation of one or more (unknown) loci is a critical step in the late stages of direct reprogramming, and that inhibition of Dnmt1 lowers this kinetic barrier, thereby facilitating the transition to pluripotency. A similar role for DNA demethylation has been reported recently during *in vivo* reprogramming in the germ line[30].

Transcription-factor-knockdown

In contrast to the other partially reprogrammed cell lines, MCV6 did not respond to AZA treatment (Supplementary Fig. 7). We also noted previously that MCV6 cells never show spontaneous appearance of GFP-positive colonies. We hypothesized that expression of one or more lineage-specifying transcription factor may have stabilized these cells in a more differentiated state than MCV8 or BIV1.

To test this hypothesis, we studied our genome-wide maps and identified lineage-specifying transcription factors that are expressed at low or undetectable levels in MCV8 or iPS cell populations. We transfected MCV6 cells with siRNAs against four transcription factors with >5-fold higher expression in MCV6 than in MCV8 (*Gata6*, *Pax7*, *Pax3* and *Sox9*). This resulted in no significant response. However, when transfection of siRNA targeting any one of the factors was followed by treatment with AZA for 48 h, GFP-positive cells appeared at a significant frequency in all examined populations (16 independent transfections; Fig. 5 and Supplementary Fig. 15). For example, targeting the primitive endoderm marker *Gata6* (ref. 31) generated ~2% GFP-positive cells within one passage of the subsequent AZA treatment. In contrast, no GFP-positive cells appeared in populations transfected with negative control siRNAs, or siRNAs targeted against transcription factors not expressed in MCV6 (*Zic1* and *Meox2*) or against *Dnmt1* (7 control populations; $P < 4 \times 10^{-4}$; Mann–Whitney U-test).

We conclude that re-activation or incomplete repression of lineage-specifying transcription factors during the reprogramming process blocks activation of the endogenous pluripotency regulatory network in MCV6. Transient silencing of one or more of these factors, combined with inhibition of Dnmt1, seems to shift the regulatory balance towards the pluripotent state, which may then be stabilized by autoregulatory feedback[11].

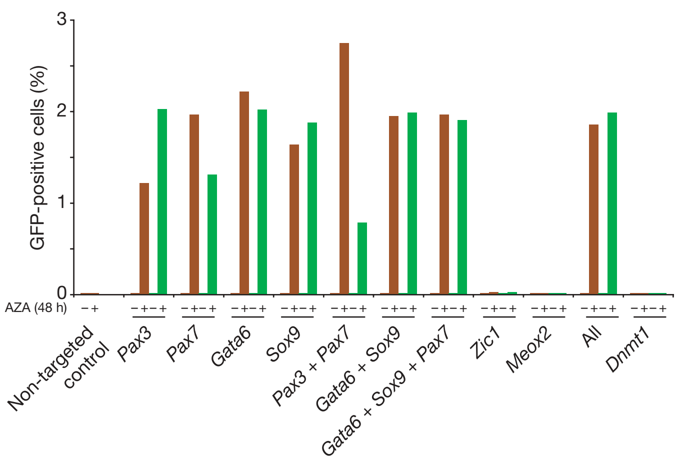


Figure 5 | Transcription factor knockdown facilitates reprogramming. MCV6 cells were plated onto 24-well dishes and transfected with siRNAs targeting expressed (*Pax7*, *Pax3*, *Gata6*, *Sox9*) or non-expressed (*Zic1*, *Meox2*) transcription factors. One plate was kept in embryonic stem cell medium and the second was exposed to AZA for 48 h. Two independent siRNA sequences were used for duplicate experiments (red and green). FACS analysis was performed 48 h after AZA treatment (96 h after transfection) without passaging. The transfection efficiency was estimated as ~20% using Cy3-coupled GADPH control siRNA.

Discussion

Several insights emerge from our integrative genomic analyses. First, the Oct4/Sox2/Klf4/c-Myc-based reprogramming process appears to be fairly general, with two independent strategies (constititutive retrovirus or inducible lentivirus) and two distinct cell types (MEFs and B lymphocytes) yielding similar immediate responses, partially reprogrammed states and a similar mechanism for the final transition to pluripotency. Second, cells may fail to reprogramme successfully for several apparent reasons: the cells may induce anti-proliferative genes in response to proliferative stress; they may inappropriately activate or fail to repress endogenous or ectopic transcription factors, and become 'trapped' in differentiated states; and they may fail to reactivate hypermethylated pluripotency genes. Third, complete reprogramming can be facilitated by direct intervention against these failure modes, such as transient inhibition of Dnmt1 and expressed transcription factors.

We expect that further characterization of intermediate states and alternative small molecule treatments will yield critical insights that will help facilitate the desired transitions, making reprogramming efficient and safe for use in regenerative medicine. More generally, our data are consistent with a model of development in which cellular states are defined by transcription factors and stabilized by epigenetic remodelling. Integrative gene expression and epigenomic profiling provides a powerful tool for defining and guiding directed transitions between these states.

Note added in proof: The work by A.M. *et al.* cited in the text as unpublished observations has now been accepted for publication[32].

METHODS SUMMARY

Embryonic stem and iPS cells were cultivated on irradiated MEFs. MEFs were infected for 16–20 h with the Moloney-based retroviral vector pLIB (Clontech) containing the complementary DNAs of *Oct4*, *Sox2*, *Klf4* and c-*Myc*. Cell lines containing the inducible lentiviruses and a ROSA26-targeted M2rtTA were induced with 2 μg ml^{-1} of Dox. AZA treatment was performed for 48 h or as indicated at a concentration of 0.5 mM.

Bisulphite treatment was performed with the Qiagen EpiTect Kit. For chromatin immunoprecipitation, cells were harvested and cross-linked with formaldehyde (final concentration 1%) for 10 min at 37 °C, were washed twice with cold PBS (plus protease inhibitors), frozen and kept at –80 °C. Chromatin immunoprecipitation, library construction, sequencing, identification of enriched intervals and chromatin state classification were performed as described previously[22]. RNA was isolated using Trizol followed by a second round of purification using RNeasy columns (Qiagen). RNA was then processed and analysed as described elsewhere[22].

Reverse transfections were performed in 24-well dishes according to manufacturer's instructions using the siPORT NeoFX transfection agent (Ambion) and Silencer Select (Ambion/ABI) siRNAs for the respective targets.

Fluorescently conjugated antibodies were used for FACS analysis and cell sorting. Cell sorting was performed by using FACS-Aria (BD-Biosciences), and consistently achieved cell sorting purity of >97%. For determining GFP-positive cell numbers by FACS, we counted >50,000 cells.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 March; accepted 8 May 2008.
Published online 28 May 2008.

1. Aoi, T. *et al.* Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science.* doi:10.1126/science.1154884 (2008).

2. Maherali N. . *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cells* **1,** 55–77 (2007).
3. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature Biotechnol.* **26,** 101–106 (2008).
4. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–317 (2007).
5. Park, I. H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451,** 141–146 (2008).
6. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–872 (2007).
7. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
8. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–1920 (2007).
9. Wernig, M. *et al. In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–324 (2007).
10. Yamanaka, S. Strategies and new developments in the generation of patient-specific pluripotent stem cells. *Cell Stem Cells* **1,** 39–49 (2007).
11. Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132,** 567–582 (2008).
12. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25,** 1177–1181 (2007).
13. Brambrink, T. *et al.* Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2,** 151–159 (2008).
14. Stadtfeld, M. *et al.* Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).
15. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
16. Adhikary, S. & Eilers, M. Transcriptional regulation and transformation by Myc proteins. *Nature Rev. Mol. Cell Biol.* **6,** 635–645 (2005).
17. Rowland, B. D. & Peeper, D. S. KLF4, p21 and context-dependent opposing forces in cancer. *Nature Rev. Cancer* **6,** 11–23 (2006).
18. Gregory, M. A., Qi, Y. & Hann S. R.. The ARF tumor suppressor: keeping Myc on a leash. *Cell Cycle* **4,** 249–252 (2005).
19. Rideout, W. M. III *et al.* Generation of mice from wild-type and targeted ES cells by nuclear cloning. *Nature Genet.* **24,** 109–110 (2000).
20. Lowry, W. E. *et al.* Generation of human induced pluripotent stem cells from dermal fibroblasts. *Proc. Natl Acad. Sci. USA* **105,** 2883–2888 (2008).
21. Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nature Rev. Genet.* **9,** 115–128 (2008).
22. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
23. Imamura, M. *et al.* Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *BMC Dev. Biol.* **6,** 34 (2006).
24. Silva, J. & Smith, A. Capturing pluripotency. *Cell* **132,** 532–536 (2008).
25. Lim, L. S. *et al.* Zic3 is required for maintenance of pluripotency in embryonic stem cells. *Mol. Biol. Cell* **18,** 1348–1358 (2007).
26. Bernstein, B. E., Meissner, A. & Lander, E. S. The mammalian epigenome. *Cell* **128,** 669–681 (2007).
27. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nature Genet.* **27,** 31–39 (2001).
28. Lei, H. *et al. De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122,** 3195–3205 (1996).
29. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33,** 5868–5877 (2005).
30. Hajkova, P. *et al.* Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452,** 877–881 (2008).
31. Singh, A. M. *et al.* A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25,** 2534–2542 (2007).
32. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* (in the press).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of the Broad Institute Genome Sequencing Platform, Genetic Analysis Platform and RNAi Platform for assistance with reagents and data generation. This research was supported by funds from the National Institutes of Health, the National Human Genome Research Institute, the National Cancer Institute, and the Broad Institute of MIT and Harvard.

Author Information All analysed data sets can be obtained from http://www.broad.mit.edu/seq_platform/chip/. Microarray and sequence data have been submitted to the NCBI GEO database under accession numbers GSE10871 and GSE11074, respectively. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.M. (alex@broad.mit.edu).

METHODS

Viral infections and cell lines. MEFs used to derive primary iPS cell lines by infections with inducible lentiviruses were harvested at 13.5 days post coitum from F_1 matings between ROSA26–M2rtTA mice[33] and Nanog–GFP mice[13]. Secondary Nanog–GFP MEFs were isolated using neomycin selection. Lentiviral preparation and infection with Dox-inducible lentiviruses encoding *Oct4*, *Klf4*, c-*Myc* and *Sox2* cDNA driven by the tetracycline operator (TetO) and a minimal cytomegalovirus (CMV) promoter were described previously[13]. MCV6 and MCV8 were generated by retroviral infection of Oct4–GFP MEFs as described previously[12].

Cell culture. Infected MEFs or secondary inducible MEFs[15] were cultured and expanded in standard embryonic stem medium and conditions[12]. Culture and viral induction were performed as described[13,15] and BIV1 was obtained as a stable line and grown under regular embryonic stem cell conditions in the presence or absence of 2 μg ml^{-1} Dox. AZA treatment was performed for 48 h or as indicated at a concentration of 0.5 mM. Higher doses showed similar effects but increased toxicity.

Expression profiling. RNA was isolated using Trizol followed by a second round of purification using RNeasy Columns (Qiagen). RNA was then processed and analysed as described elsewhere[22]. Absolute expression values were Robust Multi-Array (RMA)-normalized, truncated to absolute intensity values ≥20, and visualized using GenePattern (http://www.broad.mit.edu/cancer/software/genepattern/).

Chromatin immunoprecipitation and Illumina/Solexa sequencing. Cells were harvested and cross-linked with formaldehyde (final concentration 1%) for 10 min at 37 °C. They were washed twice with cold PBS (plus protease inhibitors), frozen and kept at –80 °C. Chromatin immunoprecipitation, library construction, sequencing, identification of enriched intervals and chromatin state classification were performed as described previously[22].

Bisulphite sequencing and COBRA. Genomic DNA was isolated and bisulphite conversion was performed in a thermocycler using the Qiagen EpiTect Kit according to manufacturer's instructions with two additional cycles (5 min at 99 °C and 3 h at 60 °C) at the end. When using 2 μg genomic DNA as starting material, converted DNA was eluted in 40 μl elution buffer (Qiagen) and 2 μl were used and amplified with previously described primer sets[23] and the following additional primer pairs (CyctF: GAAGGATTAAATAGATGTATAAGA AAATAT; CyctR: AAACCCTAATTATAAACAAATACAAC; Sox2F: GGTTTA GGAAAAGGTTGGGAATA; Sox2R: AACCAAAATAAAACAAAACCCATAA). PCR was performed in 25-μl reactions using EpiTect MSP Kit (Qiagen) mastermix according to the manufacturer's instructions with a 45 s annealing step at 50 °C (35 cycles). PCR products were gel-purified, TOPO-cloned (Invitrogen) and sequenced. COBRA for *Dppa5*, *Nanog* and *Utf1* was performed using 15 μl of the gel-purified DNA. *Dppa5* was digested for 4 h at 65 °C with Taq1 (TCGA). *Nanog* and *Utf1* were digested with HpyCHIV (ACGT) for 4 h at 37 °C. Digested products were run on 2% agarose gels.

Knockdown of transcription factors and *Dnmt1*. Reverse transfections were performed in 24-well dishes according to the manufacturer's instructions using the siPORT NeoFX transfection agent (Ambion). The following Silencer Select (Ambion/ABI) siRNAs were used (the sequence shown is the sense strand): negative control siRNA (4390843: sequence not provided), positive control Cy3 *GAPDH* siRNA (AM4649: sequence not provided), *Pax3* siRNA (s71259, GCCCACGUCUAUUCCACAA; s71260, GCUCCGAUAUUGACUCUGA), *Pax7* siRNA (s71271, CCCUCAGUGAGUUCGAUUA; s71272, CCACAUCC GUCACAAGAUA), *Gata6* siRNA (s66489, CAAAAAUACUUCUCCUUCU; s66490, CCUCUGCACGCUUUCCCUA), *Sox9* siRNA (s74192, AGACU CACAUCUCUCCUAA; s74193, AAGUUGAUCUGAAGCGAGA), *Meox2* siRNA (s69792, GCAGUGAAUCUAGACCUCA; s69793, GCCCAUCAU AAUUAUCUGA), *Zic1* siRNA (s76384, CAAAAAGUCGUGCAACAAA; s76385, GGGACUUUCUGUUCCGCA) and *Dnmt1* siRNA (s65071, GGU AGAGAGUUACGACGAA; s65072, CAACGGAUCCUAUCACACU). *Dnmt1* was stably knocked down using five independent shRNAs from the RNA interference consortium (TRC; http://www.broad.mit.edu/genome_bio/trc/). shRNA1 (TRCN0000039024; target: GCTGACACTAAGCTGTTTGTA), shRNA2 (TRCN0000039025; target: GCCTTTACTTTCAACATCAAA), shRNA3 (TRCN0000039026; target: CCGCACTTACTCCAAGTTCAA), shRNA4 (TRCN0000039027; target: CCCGAAGATCAACTCACCAAA) and shRNA5 (TRCN0000039028; target: GCAAAGAGTATGAGCCAATAT). MCV8 cells were infected overnight and selected in puromycin (final, 2 μg ml^{-1}) for 48 h.

Quantitative RT–PCR. Total RNA was isolated using RNeasy Kit (Qiagen). Three micrograms of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA-Free RNA kit (Zymo Research). Retroviral expression levels were determined as described previously[9]. For inducible lentiviral expression, 1 μg of DNase I-treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately resuspended in 100 μl of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction in an ABI Prism 7000 (Applied Biosystems) with Platinum SYBR green qPCR SuperMix-UDG with ROX (Invitrogen). Primers used for amplification were as follows: c-*Myc*: F, 5′-ACCTAACTCGAGGAGGAGCTGG-3′, and R, 5′-TCCACATAGCGTAAA AGGAGC-3′; *Klf4*: F, 5′-ACACTGTCTTCCCACGAGGG-3′, and R, 5′-GGCATTAAAGCAGCGTATCCA-3′; *Sox2*: F, 5′-CATTAACGGCACACTG CCC-3′, and R, 5′-GGCATTAAAGCAGCGTATCCA-3′; *Oct4*: F, 5′-AGCCTGGCCTGTCTGTCACTC-3′, and R, 5′-GGCATTAAAGCAGC GTATCCA-3′. To ensure equal loading of cDNA into qRT–PCR reactions, *GAPDH* messenger RNA was amplified using the following primers: F, 5′-TTCACCACCATGGAGAAGGC-3′, and R, 5′-CCCTTTTGGCTCCACCCT-3′. Data were extracted from the linear range of amplification. All graphs of qRT–PCR data shown represent samples of RNA that were DNase-treated, reverse transcribed, and amplified in parallel to avoid variation inherent in these procedures.

Flow cytometry analysis and cell sorting. The following fluorescently conjugated antibodies (PE, FITC, Cy-Chrome or APC-labelled) were used for FACS analysis and cell sorting: anti-SSEA1 (RnD Systems), anti-Igκ, anti-Igλ1,2,3, anti-CD19, anti-B220, anti-sIgM and anti-sIgD (all obtained from BD-Biosciences). Cell sorting was performed by using FACS-Aria (BD-Biosciences), and consistently achieved cell sorting purity of >97%. For determining GFP-positive cell numbers by FACS, we counted >50,000 cells.

33. Beard, C. *et al.* Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* 44, 23–28 (2006).

Vol 454 | 3 July 2008 | doi:10.1038/nature07086

ARTICLES

SMAD proteins control DROSHA-mediated microRNA maturation

Brandi N. Davis[1,2], Aaron C. Hilyard[2], Giorgio Lagna[2] & Akiko Hata[1,2]

MicroRNAs (miRNAs) are small non-coding RNAs that participate in the spatiotemporal regulation of messenger RNA and protein synthesis. Aberrant miRNA expression leads to developmental abnormalities and diseases, such as cardiovascular disorders and cancer; however, the stimuli and processes regulating miRNA biogenesis are largely unknown. The transforming growth factor β (TGF-β) and bone morphogenetic protein (BMP) family of growth factors orchestrates fundamental biological processes in development and in the homeostasis of adult tissues, including the vasculature. Here we show that induction of a contractile phenotype in human vascular smooth muscle cells by TGF-β and BMPs is mediated by miR-21. miR-21 downregulates PDCD4 (programmed cell death 4), which in turn acts as a negative regulator of smooth muscle contractile genes. Surprisingly, TGF-β and BMP signalling promotes a rapid increase in expression of mature miR-21 through a post-transcriptional step, promoting the processing of primary transcripts of miR-21 (pri-miR-21) into precursor miR-21 (pre-miR-21) by the DROSHA (also known as RNASEN) complex. TGF-β- and BMP-specific SMAD signal transducers are recruited to pri-miR-21 in a complex with the RNA helicase p68 (also known as DDX5), a component of the DROSHA microprocessor complex. The shared cofactor SMAD4 is not required for this process. Thus, regulation of miRNA biogenesis by ligand-specific SMAD proteins is critical for control of the vascular smooth muscle cell phenotype and potentially for SMAD4-independent responses mediated by the TGF-β and BMP signalling pathways.

Mutations in molecules of the TGF-β or BMP signalling pathways are found among patients with vascular disorders, indicating the essential role of TGF-β or BMP pathways in vascular homeostasis[1,2]. Both TGF-βs and BMPs are known to be critical modulators of the vascular smooth muscle cell (VSMC) phenotype[3–5]. Inhibition of TGF-β or BMP signalling in VSMCs decreases the expression of VSMC-specific genes and transforms VSMCs from a fully differentiated or 'contractile' phenotype to a dedifferentiated or 'synthetic' state[4–6].

miR-21 modulates smooth muscle phenotype

We investigated the involvement of miRNAs in the TGF-β-family-mediated modulation of the VSMC phenotype by cloning and comparing the relative abundance of miRNAs expressed in vehicle- and BMP4-treated human primary pulmonary artery smooth muscle cells (PASMCs; Supplementary Fig. 1). The expression level of a selected group of miRNAs was then directly measured by quantitative polymerase chain reaction with reverse transcription (qRT–PCR) after 24 h of BMP4 stimulation (Fig. 1a): mature miR-21 and miR-199a showed a significant increase in expression (5.7-fold and 2.1-fold, respectively) in the presence of BMP4 ($P < 0.05$). miR-21 was comparably induced by three BMP ligands that stimulate VSMC differentiation (BMP2, BMP4 and BMP7)[5] (Supplementary Fig. 2). Thus, a subset of miRNAs is induced by BMP signalling in VSMCs. High expression of miR-21 has also been observed in the vascular wall of balloon-injured rat carotid arteries—an *in vivo* model recapitulating smooth muscle phenotype switch[7].

The function of miRNAs was tested by transfecting PASMCs with 'anti-miRs': 2′-*O*-methyl-modified RNA oligonucleotides complementary to individual miRNA sequences[8]. Anti-miR-21 specifically decreased mature miR-21 expression (Supplementary Fig. 3) and effectively reduced both basal and BMP4-induced expression of the smooth muscle cell (SMC) markers smooth muscle α-actin (*SMA*, also known as *ACTA2*) and calponin (also known as *CNN1*) (Fig. 1b and Supplementary Fig. 4a, b), suggesting that miR-21 is necessary for SMC-specific gene expression. Downregulation of different miRNAs showed specific effects: targeting miR-125a and miR-125b inhibited SMC markers (Fig. 1b and Supplementary Fig. 4a, b), whereas depletion of miR-221 and miR-15b stimulated basal SMA expression (Fig. 1b and Supplementary Fig. 4b). Anti-miR-21 also decreased SMA in pluripotent mouse C3H10T1/2 (10T1/2) cells treated with BMP4 (Supplementary Fig. 4c). In gain-of-function experiments, forced expression of miR-21 by infection with an adenoviral *miR-21* construct (Ad-miR-21)[9] increased SMA protein and mRNA levels in PASMCs (Fig. 1c and Supplementary Fig. 5). Thus, miR-21 is a critical mediator of SMC differentiation by BMP signalling.

PDCD4 is a critical target of miR-21 in vascular smooth muscle

Because miR-21 has been shown to target the tumour suppressor gene *PDCD4* and to downregulate its expression in cancer cells[10–12], we asked whether PDCD4 mediates the effect of miR-21 in SMCs. Forced expression of miR-21 and reduction of miR-21 by anti-miR-21 in PASMCs decreased and increased *PDCD4* mRNA expression, respectively (Supplementary Fig. 6a, b), confirming that *PDCD4* is a miR-21 target. BMP4 treatment reduced *PDCD4* (~30%; Supplementary Fig. 6a, b) and anti-miR-21 abolished this effect (Supplementary Fig. 6b), suggesting that *PDCD4* is negatively regulated by BMP4 as a result of miR-21 induction. We next examined whether modulation of *PDCD4* expression in PASMCs affects SMC marker expression. Transfection of a human *PDCD4* expression construct, which includes a miR-21 target sequence in its 3′ untranslated region (UTR)[11] (Supplementary Fig. 6c), increased the expression of human *PDCD4* in 10T1/2 cells (Fig. 1d, right panel) and inhibited basal and BMP4-induced expression of the SMC markers *Sma*, calponin and *Sm22α* (also known as *Tagln*), but not of *Id3*, a gene directly

[1]Department of Biochemistry, Tufts University School of Medicine, [2]Molecular Cardiology Research Institute, Tufts Medical Center, Boston, Massachusetts 02111, USA.

regulated by BMP4 (ref. 13), indicating that PDCD4 represses specifically SMC genes, and has no effect on BMP signalling in general (Fig. 1d, left panel). BMP4 treatment still significantly augmented SMC gene expression and decreased ectopic *PDCD4* mRNA ($P < 0.001$), presumably through the 3′ UTR miR-21 target site (Fig. 1d). Conversely, *PDCD4* knockdown (~60%) by siRNA (*PDCD4*-siRNA) in PASMCs increased the basal expression of SMA, calponin and SM22α approximately twofold (Fig. 1e). BMP4 failed to induce SMA above the basal level when PDCD4 was depleted in the cell (Fig. 1e), whereas the levels of calponin and SM22α were still induced by BMP4 treatment, suggesting that BMP4 induces calponin and SM22α in part through a PDCD4-independent mechanism[5,14] (Fig. 1e). In conclusion, *PDCD4* is a functional target of miR-21 involved in the BMP-mediated induction of SMC markers in VSMCs.

TGF-β and BMP signalling promote miR-21 processing

TGF-β, another inducer of the contractile phenotype[3–5], stimulated the expression of both miR-21 and miR-199a to a level comparable to BMP4 (Fig. 2a) with similarly fast kinetics (2 h; Supplementary Fig. 7), indicating that TGF-β and BMPs both support a contractile phenotype by means of an increase of miR-21.

The biogenesis of miRNAs initiates with the transcription of the miRNA gene and proceeds with the cropping of the primary transcript (pri-miRNA) into a hairpin intermediate (pre-miRNA) by the nuclear ~650 kilodalton microprocessor complex, comprised in humans of the RNase III DROSHA[15], the DiGeorge syndrome critical region gene 8 (DGCR8)[16,17], and the DEAD box RNA helicases p68 and p72 (also known as DDX17)[18]. The pre-miRNA is then exported from the nucleus and processed into a ~22-nucleotide miRNA duplex by the cytoplasmic RNase III DICER[19–21]. Regulation of miRNA expression has been documented at the transcriptional level, but little is known about the stimuli and molecules regulating post-transcriptional processing[19,21–26]. BMPs and TGF-βs control gene expression through the SMAD proteins, which embody the qualities of both signal transducers and transcriptional modulators[27,28], but are not known to affect RNA processing. Therefore, we examined the accumulation of primary *miR-21* gene transcripts (pri-miR-21), pre-miR-21 and mature miR-21 on BMP or TGF-β treatment in an expression time course (Fig. 2b), expecting to find a transcriptional induction of pri-miR-21 transcripts in response to factor

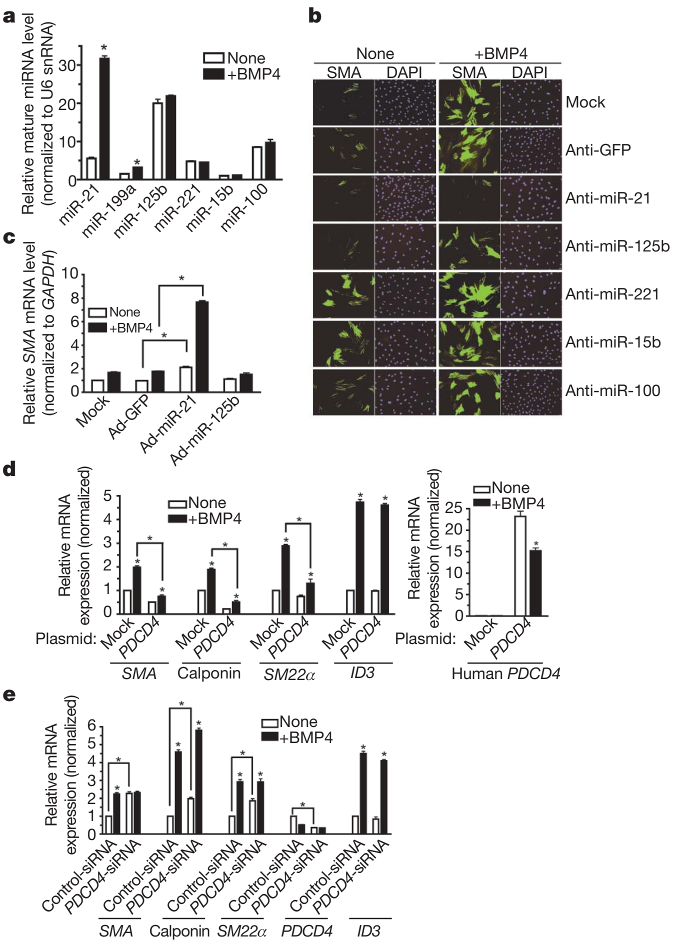


Figure 1 | miR-21 is critical for the modulation of the VSMC phenotype by BMP. a, The level of expression of miRNAs normalized to U6 small nuclear RNA (snRNA) in PASMCs treated with BMP4 for 24 h (*$P < 0.05$, $n = 4$). **b**, PASMCs transfected with antisense RNA oligonucleotides against different miRNAs or GFP (control). After BMP4 treatment (48 h), cells were stained with anti-SMA antibody (green) and 4,6-diamidino-2-phenylindole (DAPI; blue). **c**, PASMCs were infected with an adenovirus carrying CMV-driven GFP (control; Ad-GFP), miR-21 (Ad-miR-21) or miR-125b (Ad-miR-125b). The *SMA* mRNA level was measured after BMP4 treatment (48 h) (*$P < 0.05$, $n = 4$). **d**, 10T1/2 cells were transfected with vector (mock) or a human *PDCD4* cDNA construct, followed by BMP4 treatment (24 h). Expression of *Sma*, calponin, *Sm22α*, *Id3* or human *PDCD4* relative to *GAPDH* mRNA is shown (*$P < 0.001$, $n = 3$). **e**, PASMCs transfected with control siRNA (Control-siRNA) or siRNA for *PDCD4* (*PDCD4*-siRNA). Relative mRNA expression is shown as in **d**. Error bars represent s.e.m.

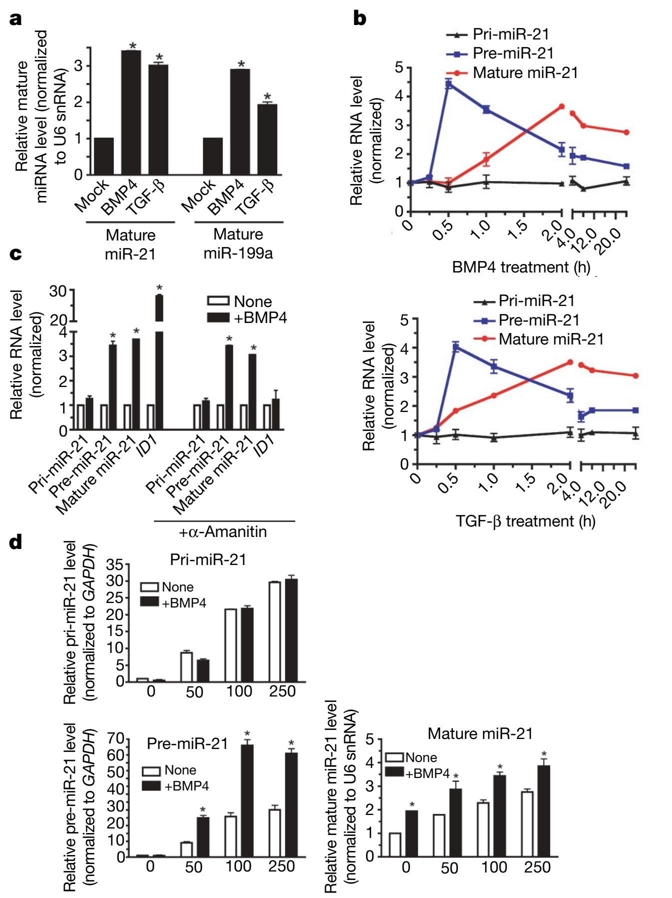


Figure 2 | Post-transcriptional regulation of miR-21 biosynthesis by TGF-β. a, Expression of mature miR-21 and miR-199a normalized to U6 snRNA in PASMCs stimulated with BMP4 or TGF-β (24 h; *$P < 0.05$; $n = 3$). **b**, Time course of pri-miR-21, pre-miR-21 or mature miR-21 expression in PASMCs on stimulation with BMP4 (upper panel) or TGF-β (lower panel). **c**, PASMCs pretreated with α-amanitin were stimulated with BMP4 (5 h). Expression of pri-miR-21, pre-miR-21 and mature miR-21 or *ID1* is shown (*$P < 0.05$; $n = 3$). **d**, Relative expression of pri-miR-21, pre-miR-21 and mature miR-21 derived from increasing amounts of human miR-21 expression construct (pCMV-miR-21) transfected into 10T1/2 cells (*$P < 0.05$; $n = 3$). Error bars represent s.e.m.

stimulation[29]. However, although we observed induction of mature miR-21 and pre-miR-21 2 h after BMP4, BMP2 or TGF-β treatment, we detected no significant change in the expression of pri-miR-21 ($P < 0.05$) after factor addition (Fig. 2b and Supplementary Fig. 8a), suggesting that induction of miR-21 by BMP4, BMP2 or TGF-β occurs at a post-transcriptional step. Likewise, BMP4-mediated induction of both pre-miR-21 and mature miR-21 was resistant to inhibition of RNA polymerase II by α-amanitin, whereas induction of the BMP4 transcriptional target gene *ID1* (ref. 30) was abolished (Fig. 2c). Furthermore, a luciferase reporter construct containing the *miR-21* gene promoter was not activated by BMP4 or TGF-β treatment, whereas it was induced by its known regulator STAT3 (Supplementary Fig. 9)[31].

A dose-dependent increase of all three forms of miR-21 was observed on transfection in mouse 10T1/2 cells of pCMV-miR-21, a plasmid in which human pri-miR-21 is transcribed from the cytomegalovirus (CMV) promoter[32] (Fig. 2d), indicating an expression level proportional to the episomal DNA copies. However, BMP4 could further induce pre-miR-21 and mature miR-21, but not pri-miR-21 (Fig. 2d), indicating that the miR-21 promoter or genomic locus is not required for post-transcriptional induction of miR-21 by BMP4. The plasmid-derived miR-21 induced by BMP4 was functional, because it repressed a miR-21 sensor construct containing complementary binding sites for the miR-21 sequence at the 3′ UTR of a luciferase reporter gene (Supplementary Fig. 8b). Furthermore, expression of CMV-transcribed *miR-21* induced SMA mRNA and protein in 10T1/2 cells in a dose-dependent manner, and was further increased by BMP4 stimulation (Supplementary Fig. 10a, b). Thus, the BMP4 pathway promotes the expression of precursor and functional mature miR-21 through a post-transcriptional, genome-independent mechanism.

SMADs interact with the RNA helicase p68

We investigated the molecular pathway leading to miR-21 induction by RNA interference (RNAi) knockdown (~80%, *SMAD*-siRNA) of the BMP-specific receptor-specific SMAD proteins (R-SMADs) expressed in PASMCs (SMAD1 and SMAD5; Fig. 3a, bottom panel, and Supplementary Fig. 11). *SMAD*-siRNA abolished BMP4 induction of both pre-miR-21 and mature miR-21, whereas the level of expression of pri-miR-21 was not affected (Fig. 3a, top panel). Induction of *SMA* and of the BMP transcriptional target *ID3* was also inhibited by SMAD1 and SMAD5 depletion, as expected (Fig. 3a, bottom panel). Therefore, R-SMADs are required for pre-miR-21 stimulation by BMP4.

We postulated that the requirement of SMADs for pre-miR-21 induction might entail a direct involvement of SMADs in the DROSHA microprocessor complex on the basis of a previous report of a constitutive interaction between the carboxy-terminal MH2 domain of SMAD1 and the RNA helicase p68 (ref. 33), a critical subunit of the DROSHA microprocessor complex[18]. To examine whether p68 is involved in the regulation of miR-21 expression by BMP4, *p68* was downregulated in PASMCs by siRNA (~70%, Supplementary Fig. 12). Expression of pri-miR-21 and the BMP4 target gene *ID3* (ref. 13) did not change significantly (Supplementary Fig. 12b), but induction of pre-miR-21 and mature miR-21 by BMP4 was completely abolished (Fig. 3b), indicating an essential role of p68 in the TGF-β and BMP-regulated synthesis of pre-miR-21.

We found that the interaction between exogenous SMAD1 and p68 is BMP4-inducible in Cos7 cells (Supplementary Fig. 13a). In *in vitro* glutathione *S*-transferase (GST) pull-down, p68 interacts both with BMP-specific SMAD1 or SMAD5 and with TGF-β-specific SMAD3, suggesting that induction of pre-miR-21 by TGF-β may also involve an R-SMAD–p68 complex (Supplementary Fig. 14a). No interaction was observed between p68 and the cofactor SMAD4 (Supplementary Fig. 14a) or the inhibitor SMAD6 (data not shown). The interaction between R-SMADs and p68 was resistant to RNase A treatment, suggesting that R-SMADs and p68 interact in the absence of pri-miRNAs (Supplementary Fig. 15). We also confirmed that the carboxyl-terminal MH2 domain of SMAD1 is sufficient to pull down p68 (ref. 33), whereas the amino-terminal MH1 domain does not bind p68 (Supplementary Fig. 14b). Thus, by binding p68, SMAD1 may be recruited to the DROSHA microprocessor complex. Indeed, on BMP4 stimulation, endogenous SMAD1 or SMAD5 could be co-immunoprecipitated with DROSHA from PASMCs (Fig. 3c) or Cos7 extracts expressing tagged DROSHA and SMAD1 (Supplementary Fig. 13b). The interaction of R-SMADs with DROSHA was markedly reduced by RNase A treatment (Supplementary Fig. 15), suggesting that the association of R-SMADs with DROSHA, unlike the R-SMADs–p68 complex, may be facilitated by miRNA transcripts. Therefore, after ligand stimulation, SMADs associate with the DROSHA microprocessor complex by means of interaction with p68, ultimately promoting accumulation of specific pre-miRNAs.

Ligand-induced association of R-SMADs with pri-miRNAs

To test whether the R-SMAD–p68–DROSHA complex assembles specifically on pri-miR-21, we performed an RNA-chromatin

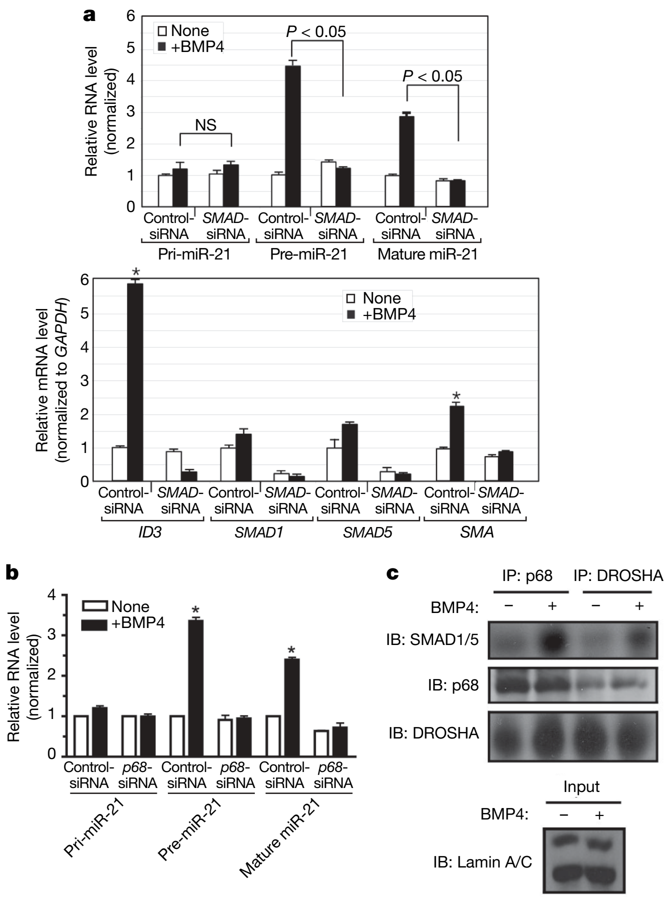


Figure 3 | Interaction of SMADs with p68, a component of the DROSHA complex. a, PASMCs were transfected with control siRNA (Control-siRNA) or a mixture of siRNAs for *SMAD1* and *SMAD5* (*SMAD*-siRNA). After BMP4 treatment (2 h), the expression of pri-miR-21, pre-miR-21 and mature miR-21 was compared (top panel). As controls, expression of *ID3*, *SMAD1*, *SMAD5* and *SMA* is shown (bottom panel). NS, not significant ($P > 0.05$). **b**, PASMCs were transfected with control siRNA (Control-siRNA) or siRNAs for *p68* (*p68*-siRNA). Expression of pri-miR-21, pre-miR-21 and mature miR-21 was examined after BMP4 treatment (2 h) ($^*P < 0.05$; $n = 3$). **c**, Nuclear extracts prepared from PASMCs treated with BMP4 (2 h) and subjected to immunoprecipitation with anti-p68, anti-DROSHA antibody, or non-specific IgG (control), followed by immunostaining with anti-SMAD1/5, anti-p68 or anti-DROSHA antibody. Nuclear extracts were immunostained with anti-lamin A/C antibody (control). IB, immunoblot; error bars represent s.e.m.

immunoprecipitation (ChIP) analysis on Cos7 cells co-transfected with pCMV-miR-21 and Flag-tagged SMAD1, SMAD3 or SMAD2. The association of SMAD1 (but not SMAD2 or SMAD3) with pri-miR-21 was induced threefold on BMP4 stimulation for 2 h (Fig. 4a and Supplementary Fig. 16a), whereas TGF-β increased binding to pri-miR-21 by SMAD2 and SMAD3, but not by SMAD1, indicating that the association between R-SMADs and pri-miR-21 is specifically regulated by ligand stimulation (Fig. 4a).

R-SMADs also interacted in a ligand-specific manner with pri-miR-21 in PASMCs (Fig. 4b), whereas p68 constitutively associated with pri-miR-21 and the recruitment of DROSHA was moderately enhanced by either TGF-β or BMP4 (Fig. 4b). Similar results were obtained for miR-199a (Fig. 4b). The significant increase ($P < 0.05$) we observed in the association of DROSHA with pri-miR-21 and pri-miR-199a (Fig. 4b) suggests that binding of SMADs to the pri-miRNA might stabilize the association between DROSHA and the pri-miRNA. We detected a constitutive association of pri-miR-214 with p68 and DROSHA, but no interaction with SMADs (Fig. 4b), confirming that pre-miR-214 is not regulated by BMP or TGF-β signals (Supplementary Fig. 17). Thus, recruitment of SMADs to the p68–DROSHA complex is pri-miRNA-specific.

A SMAD1 mutant that was non-phosphorylatable on BMP stimulation (SMAD1(3SA)) retained the ability to interact with pri-miR-21 (Supplementary Fig. 16a). Furthermore, bacterially expressed unphosphorylated GST–SMAD fusion proteins are able to interact with p68 (Supplementary Figs 14 and 15), indicating that receptor-mediated phosphorylation of R-SMADs is not essential for the association with pri-miRNA and suggesting that BMPs may affect the association between SMAD1 and pri-miRNAs primarily by controlling SMAD nuclear localization.

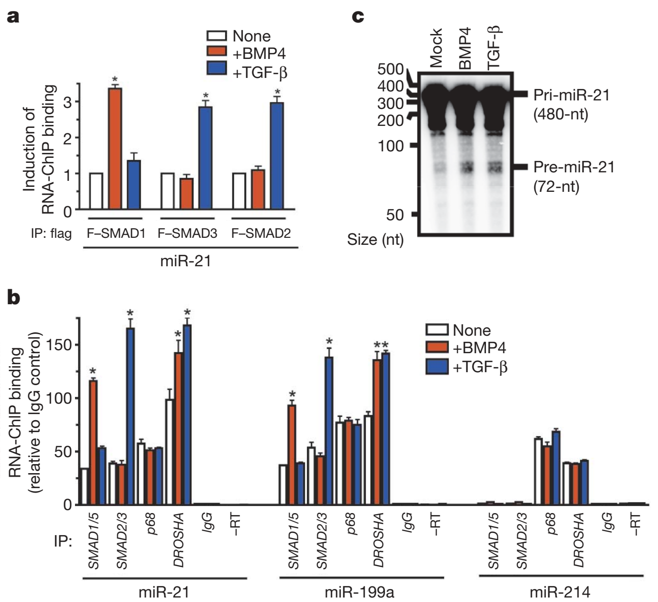


Figure 4 | Association of SMADs with pri-miRNA promotes processing by DROSHA. a, Cos7 cells transfected with pCMV-miR-21 and Flag–SMAD1, Flag–SMAD3 or Flag–SMAD2, followed by BMP4 or TGF-β treatment (2 h). RNA-ChIP performed with anti-Flag antibody or non-specific IgG (control), followed by PCR amplification with miR-21 primers (*$P < 0.05$, compared to no treatment; $n = 4$). IP, immunoprecipitation. **b**, After treatment of PASMCs with BMP4 or TGF-β (1 h), endogenous *SMAD1*/*SMAD5*, *SMAD2*/*SMAD3*, *p68* or *DROSHA* were immunoprecipitated and subjected to PCR analysis with miR-21, miR-199a or miR-214 primers. As controls, RNA samples untreated with reverse transcriptase (−RT) or immunoprecipitated with non-specific *IgG* (*IgG*) were subjected to PCR (*$P < 0.05$ compared to none; $n = 4$). **c**, *In vitro* pri-miRNA processing assay performed by incubating pri-miR-21 substrate with the nuclear extracts prepared from Cos7 cells treated with vehicle, BMP4 or TGF-β (2 h). nt, nucleotide; error bars represent s.e.m.

Pull-down experiments using partially purified GST–SMAD fusion proteins as bait confirmed that SMAD1, SMAD3 and SMAD5 can interact with pri-miR-21. Interestingly, both the MH1 and the MH2 domains of SMAD1 bound to pri-miR-21 (Supplementary Fig. 18). Because MH1 does not interact with p68 (Supplementary Fig. 14b), it is possible that MH1 interacts either with pri-miR-21 directly or with other miR-21-binding proteins.

In summary, BMPs and TGF-β stimulate the expression of a specific subset of miRNAs by inducing the formation of a complex comprising R-SMAD proteins, pri-miRNAs and subunits of the microprocessor complex such as DROSHA and p68.

Finally, we examined the possibility that ligand treatment may facilitate DROSHA-mediated production of pre-miRNA. *In vitro* pri-miRNA processing assays were performed by incubating radio-labelled pri-miR-21 substrate (480 nucleotides) with nuclear extracts from Cos7 cells treated with vehicle, BMP4 or TGF-β. Ligand treatment resulted in ~25% increase (BMP4, 28.5% ± 1.9% (mean ± s.e.m.; TGF-β, 24.2% ± 1.4%; triplicate experiments) in the production of a 72-nucleotide product corresponding to pre-miR-21, compared to incubation with extracts from mock-treated cells (Fig. 4c). This result suggests that ligand-induced association of SMADs with the DROSHA complex increases pri-miR-21 cropping into pre-miRNA.

SMAD4-independent regulation of miRNA processing

Two observations led us to speculate that SMAD4 may be dispensable for the regulation of miR-21 processing: the lack of interaction between p68 and SMAD4 (Supplementary Fig. 14a), the common SMAD required for most transcriptional responses to BMP and TGF-β signalling; and the ability of the SMAD1(3SA) mutant, which does not form a complex with SMAD4 (ref. 34), to associate with pri-miR-21 (Supplementary Fig. 16a). Transfection of an siRNA against *SMAD4* (*SMAD4*-siRNA) in PASMCs markedly reduced SMAD4 protein (~90%, Supplementary Fig. 18a) and RNA (Fig. 5a), as well as the transcriptional inducibility of the BMP target gene *ID3* (ref. 13; from 18-fold to 3-fold), as expected (Supplementary Fig. 18b). However, *SMAD4*-siRNA did not affect the induction of pre-miR-21 or mature miR-21 by BMP4 (Fig. 5a), in contrast with the result obtained from downregulation of R-SMADs (Fig. 3a). Therefore, SMAD4 is not required for the stimulation of processing of miR-21 by BMP4 in PASMCs.

Cancer cells in which the canonical TGF-β pathway is impaired, such as the *SMAD4*-negative MDA-MB-468 cells, lack the ability to transcriptionally regulate most TGF-β target genes[35,36] but retain some TGF-β responses, such as nuclear translocation of R-SMADs, increased cell migration and epithelial-to-mesenchymal transition[36–38]. We investigated whether miR-21 stimulation by TGF-β can still occur in MDA-MB-468 cells as it does in PASMCs depleted of SMAD4. A rapid induction of pre-miR-21 and mature miR-21 was observed on TGF-β stimulation in MDA-MB-468 cells, without a change in the levels of pri-miR-21 or of the SMAD4-dependent TGF-β target gene plasminogen activator inhibitor-1 (*PAI-1*, also known as *SERPINE1*) mRNA[35,36] (Fig. 5b). Similar results were obtained by BMP4 treatment of MDA-MB-468 and SMAD4-expressing breast carcinoma MCF7 cells (Supplementary Fig. 19) or by TGF-β treatment of SMAD4-positive breast carcinoma MDA-MB-231 cells (Supplementary Fig. 20). Stimulation of pri-miRNA processing by TGF-β does not necessarily lead to an increase in mature miRNA: unlike MDA-MB-468 cells (Fig. 5b), MDA-MB-231 cells display little increase of mature miR-21 after TGF-β stimulation despite strong induction of pre-miR-21 (Supplementary Fig. 20), suggesting the existence of another regulatory step of miRNA maturation after pri-miRNA cleavage by the DROSHA microprocessor. An RNA-ChIP analysis confirmed that in MDA-MB-468 cells the association of R-SMADs with the primary transcripts of miR-21 and miR-199a (but not miR-214) is ligand-inducible (Fig. 5c and Supplementary Fig. 21). Therefore, SMAD4 is not necessary for

ligand-mediated processing of pri-miRNAs, and some of the SMAD4-independent responses observed in ligand-stimulated cells may be mediated by regulation of miRNA biogenesis by the TGF-β or BMP pathways.

TGF-β increases miR-21 expression in breast carcinoma

The expression of mature miR-21 is augmented in different types of tumours and tumour-derived cell lines, including breast carcinoma MCF7, MDA-MB-231 and MDA-MB-468 cells[11,39–43]. Because TGF-β expression is often increased in cancer cells, where it promotes epithelial-to-mesenchymal transition and metastatic behaviour[44–48], we postulated that the increased levels of miR-21 may in part be caused by autocrine TGF-β signalling. A dominant-negative TGF-β type I receptor (ALK5, also known as TGFBR1)[49], which harbours a mutation in the kinase domain, was expressed in MDA-MB-468 cells to inhibit TGF-β signalling. Both the basal and the TGF-β-induced expression of pre-miR-21 were greatly reduced, whereas the pri-miR-21 level was unchanged (Fig. 5d). These results indicate that autocrine TGF-β signalling contributes to the high basal expression of miR-21 in cancer cells.

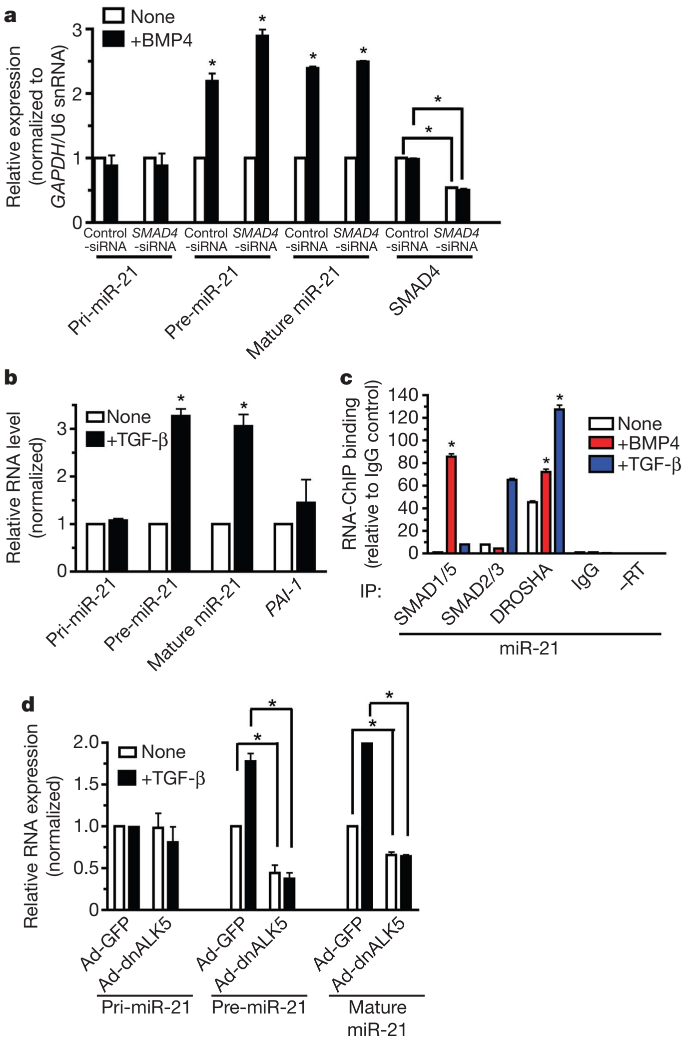


Figure 5 | SMAD4-independent mechanism of maturation of pri-miRNA. **a**, Level of expression of pri-miR-21, pre-miR-21 and mature miR-21 or *SMAD4* after treatment with BMP4 (2 h) in PASMCs transfected with control siRNA (Control-siRNA) or *SMAD4* siRNA (*SMAD4*-siRNA). **b**, Level of expression of pri-miR-21, pre-miR-21 and mature miR-21 or *PAI-1* in human *SMAD4*-negative breast carcinoma MDA-MB-468 cells stimulated with TGF-β (0.5 h) (*$P < 0.05$; $n = 3$). **c**, MDA-MB-468 cells were treated with TGF-β (1 h) before RNA-ChIP. Endogenous proteins were precipitated with anti-SMAD1/SMAD5, anti-SMAD2/SMAD3 or anti-DROSHA antibodies, followed by PCR analysis with a miR-21 primer (*$P < 0.05$, compared to none; $n = 3$). **d**, MDA-MB-468 cells were infected with adenovirus carrying dominant-negative type I TGF-β receptor (dnALK5), an inhibitor of TGF-β signalling, before TGF-β treatment (1 h). The amount of pri-miR-21, pre-miR-21 and mature miR-21 was examined (*$P < 0.05$, compared to none; $n = 3$). Error bars represent s.e.m.

Discussion

This study underscores several unexpected findings. First, the TGF-β superfamily of growth factors triggers VSMC differentiation by increasing the expression of a subset of miRNAs. This induction occurs post-transcriptionally, probably at the level of processing of primary transcripts by the DROSHA microprocessor complex. Ligand-specific SMAD proteins bind to the DROSHA microprocessor subunit p68 to facilitate pre-miRNA accumulation. Finally, we identified a previously unknown mechanism by which the TGF-β pathway may promote the metastatic and invasive potential of cancer cells through modulation of biosynthesis of oncogenic miRNAs such as miR-21, which in turn targets tumour suppressor genes *PDCD4* and tensin homologue deleted on chromosome 10 (*PTEN*)[10,11].

Open questions remain regarding, for example, the exact number and identity of pri-miRNAs regulated by SMADs and the determinants of specificity in their selection. The MH1 domain of R-SMADs binds DNA by specifically recognizing a sequence element[27,50]; we observed that the MH1 domain of SMAD1 associates with pri-miR-21 despite its inability to interact with p68: it is possible to speculate that the SMAD MH1 domain may recognize an RNA sequence or structural element, and thus provide specificity in the selection of BMP and TGF-β target miRNA. The exact role of the SMAD–p68 interaction in the DROSHA microprocessor complex also remains unsolved. Association of SMAD with the DROSHA complex is likely to contribute to various aspects of pri-miRNA processing, such as facilitating the specific recognition and stable binding of DROSHA to pri-miRNAs, increasing the RNase activity of DROSHA, directing the cleavage of pri-miRNAs to a precise sequence, or modulating the stability of pre-miRNA. In summary, our findings open new avenues to the study of TGF-β-family signalling pathways and miRNA biogenesis regulation.

METHODS SUMMARY

Cell culture. Cos7, C3H10T1/2, MDA-MB-468, MDA-MB-231 and MCF7 cells (American Type Culture Collection) were maintained in DMEM supplemented with 10% FBS (Sigma). Human primary PASMCs were purchased from Lonza (CC-2581; http://www.lonzabioscience.com/Lonza_Catnav.oid.734.prodoid.PASMC) and were maintained in Sm-GM2 media (Lonza) containing 5% FBS.

Real-time RT–PCR. Total RNA was extracted by Trizol (Invitrogen) and subjected to reverse transcription using a first-strand cDNA synthesis kit (Invitrogen) according to the manufacturer's instructions. The quantitative analysis of the change in expression levels was calculated by real-time PCR machine (iQ5, BioRad)[29]. For detection of mature miRNAs, the TaqMan MicroRNA assay kit (Applied Biosystems) was used according to the manufacturer's instructions. An average of three experiments each performed in triplicate with standard errors is presented.

RNA-ChIP. RNA-ChIP was performed as described previously[18]. An average of three experiments each performed in triplicate with standard errors is presented.

***In vitro* pri-miRNA processing assays.** The *in vitro* pri-miRNA processing assay was performed as described previously[26].

Statistical analysis. The results presented are the average of at least three experiments each performed in triplicate with standard errors. Statistical analyses were performed by ANOVA, followed by Tukey's multiple comparison test or by Student's *t* test as appropriate, using Prism 4 (GraphPAD Software Inc.). *P* values of <0.05 were considered significant and are indicated with asterisks.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 March; accepted 9 May 2008.
Published online 11 June 2008.

1. ten Dijke, P. & Arthur, H. M. Extracellular control of TGFβ signalling in vascular development and disease. *Nature Rev. Mol. Cell Biol.* 8, 857–868 (2007).

2. Morrell, N. W. Pulmonary hypertension due to BMPR2 mutation: a new paradigm for tissue remodeling? *Proc. Am. Thorac. Soc.* **3**, 680–686 (2006).
3. Owens, G. K. Regulation of differentiation of vascular smooth muscle cells. *Physiol. Rev.* **75**, 487–517 (1995).
4. Rensen, S. S. M., Doevendans, P. A. F. M. & van Eys, G. J. J. M. Regulation and characteristics of vascular smooth muscle cell phenotypic diversity. *Netherlands Heart J.* **15**, 100–108 (2007).
5. Lagna, G. *et al.* Control of phenotypic plasticity of smooth muscle cells by BMP signaling through the myocardin-related transcription factors. *J. Biol. Chem.* **282**, 37244–37255 (2007).
6. Owens, G. K., Kumar, M. S. & Wamhoff, B. R. Molecular regulation of vascular smooth muscle cell differentiation in development and disease. *Physiol. Rev.* **84**, 767–801 (2004).
7. Ji, R. *et al.* MicroRNA expression signature and antisense-mediated depletion reveal an essential role of microRNA in vascular neointimal lesion formation. *Circ. Res.* **100**, 1579–1588 (2007).
8. Esau, C. C. Inhibition of microRNA with antisense oligonucleotides. *Methods* **44**, 55–60 (2008).
9. van Rooij, E. *et al.* A signature pattern of stress-responsive microRNAs that can evoke cardiac hypertrophy and heart failure. *Proc. Natl Acad. Sci. USA* **103**, 18255–18260 (2006).
10. Asangani, I. A. *et al.* MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. *Oncogene* **27**, 2128–2136 (2008).
11. Frankel, L. B. *et al.* Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. *J. Biol. Chem.* **283**, 1026–1033 (2007).
12. Zhu, S. *et al.* MicroRNA-21 targets tumor suppressor genes in invasion and metastasis. *Cell Res.* **18**, 350–359 (2008).
13. Hollnagel, A., Oehlmann, V., Heymer, J., Ruther, U. & Nordheim, A. *Id* genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J. Biol. Chem.* **274**, 19838–19845 (1999).
14. Chan, M. C. *et al.* A novel regulatory mechanism of the bone morphogenetic protein (BMP) signaling pathway involving the carboxyl-terminal tail domain of BMP type II receptor. *Mol. Cell. Biol.* **27**, 5776–5789 (2007).
15. Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415–419 (2003).
16. Han, J. *et al.* The Drosha–DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016–3027 (2004).
17. Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its *D. melanogaster* homolog are required for miRNA biogenesis. *Curr. Biol.* **14**, 2162–2167 (2004).
18. Fukuda, T. *et al.* DEAD-box RNA helicase subunits of the Drosha complex are required for processing of rRNA and a subset of microRNAs. *Nature Cell Biol.* **9**, 604–611 (2007).
19. Kim, V. N. & Nam, J. W. Genomics of microRNA. *Trends Genet.* **22**, 165–173 (2006).
20. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nature Rev. Mol. Cell Biol.* **6**, 376–385 (2005).
21. Zhao, Y. & Srivastava, D. A developmental view of microRNA function. *Trends Biochem. Sci.* **32**, 189–197 (2007).
22. Lee, E. J. *et al.* Systematic evaluation of microRNA processing patterns in tissues, cell lines, and tumors. *RNA* **14**, 35–42 (2007).
23. Obernosterer, G., Leuschner, P. J., Alenius, M. & Martinez, J. Post-transcriptional regulation of microRNA expression. *RNA* **12**, 1161–1167 (2006).
24. Thomson, J. M. *et al.* Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev.* **20**, 2202–2207 (2006).
25. Wulczyn, F. G. *et al.* Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *FASEB J.* **21**, 415–426 (2007).
26. Guil, S. & Caceres, J. F. The multifunctional RNA-binding protein hnRNP A1 is required for processing of miR-18a. *Nature Struct. Mol. Biol.* **14**, 591–596 (2007).
27. Massague, J., Seoane, J. & Wotton, D. Smad transcription factors. *Genes Dev.* **19**, 2783–2810 (2005).
28. Schmierer, B. & Hill, C. S. TGFβ–SMAD signal transduction: molecular specificity and functional flexibility. *Nature Rev. Mol. Cell Biol.* **8**, 970–982 (2007).
29. Schmittgen, T. D. *et al.* Real-time PCR quantification of precursor and mature microRNA. *Methods* **44**, 31–38 (2008).
30. Korchynskyi, O. & ten Dijke, P. Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. *J. Biol. Chem.* **277**, 4883–4891 (2002).
31. Loffler, D. *et al.* Interleukin-6 dependent survival of multiple myeloma cells involves the Stat3-mediated induction of microRNA-21 through a highly conserved enhancer. *Blood* **110**, 1330–1333 (2007).
32. Zhu, S., Si, M. L., Wu, H. & Mo, Y. Y. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). *J. Biol. Chem.* **282**, 14328–14336 (2007).
33. Warner, D. R. *et al.* Functional interaction between Smad, CREB binding protein, and p68 RNA helicase. *Biochem. Biophys. Res. Commun.* **324**, 70–76 (2004).
34. Kretzschmar, M., Liu, F., Hata, A., Doody, J. & Massagué, J. The TGF-β mediator Smad1 is directly phosphorylated and functionally activated by the BMP receptor kinase. *Genes Dev.* **11**, 984–995 (1997).
35. Gomis, R. R. *et al.* A FoxO–Smad synexpression group in human keratinocytes. *Proc. Natl Acad. Sci. USA* **103**, 12747–12752 (2006).
36. Levy, L. & Hill, C. S. Smad4 dependency defines two classes of transforming growth factor β (TGF-β) target genes and distinguishes TGF-β-induced epithelial-mesenchymal transition from its antiproliferative and migratory responses. *Mol. Cell. Biol.* **25**, 8108–8125 (2005).
37. Giehl, K., Imamichi, Y. & Menke, A. Smad4-independent TGF-β signaling in tumor cell migration. *Cells Tissues Organs* **185**, 123–130 (2007).
38. Ijichi, H. *et al.* Smad4-independent regulation of p21/WAF1 by transforming growth factor-β. *Oncogene* **23**, 1043–1051 (2004).
39. Si, M. L. *et al.* miR-21-mediated tumor growth. *Oncogene* **26**, 2799–2803 (2007).
40. Diederichs, S. & Haber, D. A. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. *Cancer Res.* **66**, 6097–6104 (2006).
41. Volinia, S. *et al.* A microRNA expression signature of human solid tumors defines cancer gene targets. *Proc. Natl Acad. Sci. USA* **103**, 2257–2261 (2006).
42. Iorio, M. V. *et al.* MicroRNA gene expression deregulation in human breast cancer. *Cancer Res.* **65**, 7065–7070 (2005).
43. Wiemer, E. A. The role of microRNAs in cancer: no small matter. *Eur. J. Cancer* **43**, 1529–1544 (2007).
44. Bierie, B. & Moses, H. L. Tumour microenvironment: TGFβ: the molecular Jekyll and Hyde of cancer. *Nature Rev. Cancer* **6**, 506–520 (2006).
45. Arteaga, C. L. Inhibition of TGFβ signaling in cancer therapy. *Curr. Opin. Genet. Dev.* **16**, 30–37 (2006).
46. Bachman, K. E. & Park, B. H. Duel nature of TGF-β signaling: tumor suppressor vs. tumor promoter. *Curr. Opin. Oncol.* **17**, 49–54 (2005).
47. Glick, A. B. TGFβ1, back to the future: revisiting its role as a transforming growth factor. *Cancer Biol. Ther.* **3**, 276–283 (2004).
48. Massague, J. & Gomis, R. R. The logic of TGFβ signaling. *FEBS Lett.* **580**, 2811–2820 (2006).
49. Fujii, M. *et al.* Roles of bone morphogenetic protein type I receptors and Smad proteins in osteoblast and chondroblast differentiation. *Mol. Biol. Cell* **10**, 3801–3813 (1999).
50. Shi, Y. *et al.* Crystal structure of a Smad MH1 domain bound to DNA: insights on DNA-binding in TGFβ signaling. *Cell* **94**, 585–594 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M.-C. Chan and N. Neuman for critical discussion and H. Surks and M. Ivan for critical reading of the manuscript and technical advice. We also thank S. Kato, E. Olson, R. Bassel-Duby, Y.Y. Mo, K. Miyazono, B. Cochran and G.-R. Wang for sharing reagents. This work was supported by grants from the National Institute of Health HD042149 and HL082854 to A.H. and HL086572 to G.L.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.H. (akiko.hata@tufts.edu).

METHODS

RT–PCR primers. Human pri-miR-21: 5′-TTTTGTTTTGCTTGGGAGGA-3′ and 5′-AGCAGACAGTCAGGCAGGAT-3′. Human pre-miR-21: 5′-TGTCGG GTAGCTTATCAGAC-3′ and 5′-TGTCAGACAGCCCATCGACT-3′. Human *GAPDH*: 5′-ACCACAGTCCATGCCATCAC-3′ and 5′-TCCACCACCCT GTTGCTGTA-3′. Human *SMA*: 5′-CCAGCTATGTGTGAAGAAGAGG-3′ and 5′-GTGATCTCCTTCTGCATTCGGT-3′. Human *ID1*: 5′-CCCATTCTG TTTCAGCCAGT-3′ and 5′-TGTCGTAGAGCAGCACGTTT-3′. Human *ID3*: 5′-ACTCAGCTTAGCCAGGTGGA-3′ and 5′-AAGCTCCTTTTGTCGTTGGA-3′. Human *PDCD4*: 5′-TATGATGTGGAGGAGGTGGATGTGA-3′ and 5′-TAT GATGTGGAGGAGGTGGATGTGA-3′. Human *p68*: 5′-TATGATGTGGA GGAGGTGGATGTGA-3′ and 5′-TATGATGTGGAGGAGGTGGATGTGA-3′. Human calponin: 5′-GAGTGTGCAGACGGAACTTCAGCC-3′ and 5′-GTCTGTGCCCAGCTTGGGGTC-3′. Human *SM22α*: 5′-CGCGAAGTGC AGTCCAAAATCG-3′ and 5′-GGGCTGGTTCTTCTTCAATGGGC-3′.

siRNAs. Synthetic siRNAs targeting human *SMAD1*, *SMAD4* or *SMAD5* and *p68* were validated Stealth DuoPak (Invitrogen) and Stealth Select RNAi (Invitrogen), respectively. For *SMAD4*: 5′-CCUGAGUAUUGGUGUUCC AUUGCUU-3' and 5′-GCAAAGGUGUGCAGUUGGAAUGUAA-3'. For *SMAD1*: 5′-GCAACCGAGUAACUGUGUCACCAUU-3' and 5′-GGUCUG CAUCAAUCCCUACCACUAU-3'. For *SMAD5*: 5′-GCCACCUGAUGAUCA GAUGGGUCAA-3' and 5′-GCUUGGGUUUGUUGUCAAAUGUUAA-3'. For *p68*: 5′-GGAAUCUUGAUGAGCUGCCUAAAUU-3′, 5′-ACAACUGCCCGA AGCCAGUUCUAAA-3′, and 5′-GGUGCAGCAAGUAGCUGCUGAAUAAA-3′. siRNA for human *PDCD4* was described previously[11] and synthesized by Dharmacon. As a negative control, Stealth RNAi Negative Control Duplex number 1-3 (Invitrogen) or scrambled siRNA (Dharmacon) was used.

RNA-ChIP primers. Human miR-21: 5′-TTTTGTTTTGCTTGGGAGGA-3′ and 5′-AGCAGACAGTCAGGCAGGAT-3′. Human miR-199a: 5′-GCCAACC CAGTGTTCAGACTA-3′ and 5′-GCCTAACCAATGTGCAGACTA-3′. Human miR-214: 5′-GCCTAACCAATGTGCAGACTA-3′ and 5′-CTATGGTGT GAGGGCTGCTT-3′. Human TM: 5′-GCAAGCACATAGTGGAGCAA-3′ and 5′-TCAAACATCCAGGACAACCA-3′.

Antibodies. Anti-Flag epitope tag (M2, Sigma), anti-p68 (clone PAb204, Upstate), anti-SMA (clone 1A4, Sigma), anti-calponin (clone hCP, Sigma), anti-GAPDH (2E3-2E10, Abnova), anti-SMAD2/SMAD3 (06-654, Upstate), anti-SMAD1/SMAD5/SMAD8 (Calbiochem), anti-SMAD4 (H-552, Santa Cruz), anti-Myc epitope tag (clone 9E10, Tufts Antibody Core Facility), anti-Lamin-A/C (2032, Cell Signalling) and anti-DROSHA (07-717, Upstate) antibodies were used.

***In vitro* pri-miRNA processing assays.** *In vitro* pri-miRNA processing assay was performed as described previously[26]. In brief, the 480-nucleotide radiolabelled pri-miR-21 was prepared by standard *in vitro* transcription with T7 RNA polymerase in the presence of [α-^{32}P]–UTP using human miR-21 gene cloned into pGEM-3 vector as a template. Nuclear extracts were prepared from $\sim 5 \times 10^6$ Cos7 cells treated with vehicle, 400 pM TGF-β or 3 nM BMP4 for 2 h. After dialysis into reaction buffer, nuclear extracts were incubated with pri-miR-21 substrates for 90 min at 37 °C. Reaction mixtures were subjected to phenol-chroloform extraction, precipitation and 10% (w/v) denaturing gel electrophodresis, followed by autoradiography. The amount of pri-miR-21 (input) and pre-miR-21 was quantified by the phosphoimager (Typhoon9410, GE Healthcare) using ImageQuant 350 software (GE Healthcare).

miRNA and cDNA expression constructs. The pCMV-miR-21 construct and recombinant adenovirus carrying miR-21 or miR-125b (Ad-miR-21 or Ad-miR-125b) were reported previously[9,32]. In brief, the pCMV-miR-21 construct contains 480-bp human miR-21 genomic fragments cloned into a modified pCMV-Myc vector (Clontech). Ad-miR-21 and Ad-miR-125b contain 280-bp rat miR-21 and 366-bp miR-125b genomic fragments into CMV-driven adenoviral vector, respectively. To monitor the amount of pri-miR-21 and pre-miR-21 derived from the pCMV-miR-21 construct in mouse 10T1/2 cells, human-specific RT–PCR primers complementary to sequences in the miR-21 flanking region were used. Unlike pri-miR-21 or pre-miR-21, mature miR-21, which is identical in mouse and human, was detected as the sum of the endogenous and recombinant products. Human *PDCD4* and *p68* cDNA constructs were purchased from OriGene. In brief, a full-length human *PDCD4* cDNA with a 1.9 kb 3′ UTR (NM_01445), which contains miR-21 target sequence, was cloned into pCMV6 vector. The human *DROSHA* cDNA construct was purchased from Addgene. The Flag–SMAD1(3SA) construct (a gift from Massague laboratory) contains human *SMAD1* cDNA with Ser to Ala mutations at amino acids 462, 463 and 465 and was cloned into pCMV5 vector[34].

Plasmid DNA and siRNA transfection. Cos7, 10T1/2 cells or PASMCs were transfected with FuGENE6 (Roche Applied Science) for plasmid DNAs and Oligofectamine (Invitrogen) for siRNAs as described before[14].

Adenoviral infection. The recombinant adenoviruses were generated and purified by standard procedures. Infection of adenoviruses was performed at 100 multiplicity of infection. There was no detectable toxicity to the cells under these conditions.

qRT–PCR assays. For qRT–PCR assays, total RNA was extracted from cells by Trizol (Invitrogen). cDNA was synthesized from 1 μg of purified RNA by SuperScript II First-Strand cDNA synthesis system (Invitrogen) according to the manufacturer's instructions. qRT–PCR was performed with a real-time PCR machine (iQ5, BioRad). The results of qRT–PCR assays presented are an average of three independent RNA preparations. Each sample was analysed in triplicate. PCR cycling parameters were: 94 °C for 3 min, and 40 cycles of 94 °C for 15 s, 60 °C for 20 s, 72 °C for 40 s. For detection of mature miRNAs, TaqMan MicroRNA assay kit (Applied Biosystems) was used according to manufacturer's protocol. Data analysis was done by using the comparative C_T method in software by BioRad.

Luciferase assay. After transfection of the reporter construct together with LacZ plasmid as an internal control, the cells were reseeded onto 12-well plates and treated with 3 nM BMP3 or 400 pM TGF-β1 for 16–20 h in DMEM/0.2% FCS. Luciferase assays were carried out using Promega's Luciferase assay system. Luciferase activity was normalized with LacZ activity.

Anti-miRNAs. 2′-*O*-methyl modified RNA oligonucleotides complementary to miRNA (anti-miR) or GFP (control) sequence were purchased from IDT. Anti-miRs were transfected to cells at a concentration of 106 nM using Oligofectamine (Invitrogen) according to the manufacturer's directions. Anti-miR-21: 5′-GUCAACAUCAGUCUGAUAAGCUA-3′. Anti-miR-199a: 5′-GAACAGGUAG UCUGAACACUGGG-3′. Anti-miR-125b: 5′-UCACAAGUUAGGGUCUCAG GGA-3′. Anti-miR-221: 5′-GAAACCCAGCAGACAAUGUAGCU-3′. Anti-miR-15b: 5′-UGUAAACCAUGAUGUGCUGCUA-3′. Anti-miR-100: 5′-CACA AGUUCGGAUCUACGGGUU-3′. Anti-GFP: 5′-AAGGCAAGCUGACCCUGA AGU-3′.

miRNA cloning. miRNA cloning from PASMCs was performed following the protocol from the Bartel laboratory (http://web.wi.mit.edu/bartel/pub/protocols_reagents.htm). In brief, miRNAs were prepared from PASMCs treated with 3 nM BMP4 for 24 h using Trizol (Invitrogen). After linker ligation and PCR amplification, miRNA sequences were concatemerized, cloned into Topo-TA vector, and sequenced by the DNA sequencing facility at Tufts University.

RNA-ChIP. RNA-ChIP was performed as described previously[18]. In brief, PASMCs or Cos7 cells were crosslinked for 15 min with 1% formaldehyde, the cell pellet was resuspended in Buffer A (5 mM PIPES, pH 8.0, 85 mM KCl, 0.5% Nonidet P-40). After 10 min on ice, the crude nuclei fraction was isolated by centrifugation, and then suspended in Buffer B (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1). After nuclei were disrupted by sonication, the lysates were cleared and subjected to immunoprecipitation with anti-Flag, anti-SMAD1/5/8, SMAD2/3 or anti-p68 antibody, followed by stringent washing, and elution. The RNA was isolated using Trizol (Invitrogen). Pellets were resuspended in TE buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) and incubated with DNase I for 30 min at 37 °C to remove any remaining DNA. After extraction with phenol:chloroform (5:1), RNA was precipitated with ethanol and dissolved in 20 μl of water. 5 μl of RNA was used for a 20 μl cDNA synthesis reaction. Quantitative PCR reactions were then performed by real-time PCR machine (iQ5, BioRad).

GST pull-down assay. GST–SMAD fusion proteins were expressed in bacteria, followed by partial purification with GST–sepharose beads. Equal amounts of GST–SMAD fusion proteins conjugated to sepharose beads were added to nuclear extracts or total RNA. After washing the beads, proteins pulled-down with the beads were separated by SDS–PAGE, followed by immunoblotting or RT–PCR analysis. For RNase A treatment, 250 $\mu g\,ml^{-1}$ RNase A (New England Biolab) were added to nuclear extracts 30 min before addition of GST–SMAD fusion proteins and throughout the pull-down assay.

Immunoprecipitation/immunoblot assay. Cells were lysed in TNE buffer (1% Nonidet P-40, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA, 150 mM NaCl). Total cell lysates or proteins immunoprecipitated with antibodies were separated by SDS–PAGE, transferred to PVDF membranes (Millipore), immunoblotted with antibodies, and visualized using an enhanced chemiluminescence detection system (Amersham Biosciences).

Immunofluorescence staining. PASMCs or 10T1/2 cells were fixed and permeabilized in a 50% acetone-50% methanol solution and subjected to staining using anti-SMA or anti-calponin antibody conjugated with fluorescein isothiocyanate (FITC) and nuclear staining with DAPI (Invitrogen).

CORRIGENDUM

doi:10.1038/nature07086

SMAD proteins control DROSHA-mediated microRNA maturation

Brandi N. Davis, Aaron C. Hilyard, Giorgio Lagna & Akiko Hata

Nature 454, 56–61 (2008)

In the RNA-ChIP primers section of the Methods, human miR-214 was incorrectly listed. Human miR-214 should be 5′-CTGCTTTC TTTCAATGGCTGGTTGT-3′ and 5′-CTGATTGTATCTGTCTAT GAGCAAA-3′.

LETTERS

Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock

John D. Richardson[1,2], Justin C. Kasper[3], Chi Wang[2], John W. Belcher[1] & Alan J. Lazarus[1]

The solar wind blows outward from the Sun and forms a bubble of solar material in the interstellar medium. The termination shock occurs where the solar wind changes from being supersonic (with respect to the surrounding interstellar medium) to being subsonic. The shock was crossed by Voyager 1 at a heliocentric radius of 94 AU (1 AU is the Earth–Sun distance) in December 2004 (refs 1–3). The Voyager 2 plasma experiment observed a decrease in solar wind speed commencing on about 9 June 2007, which culminated in several crossings of the termination shock between 30 August and 1 September 2007 (refs 4–7). Since then, Voyager 2 has remained in the heliosheath, the region of shocked solar wind. Here we report observations of plasma at and near the termination shock and in the heliosheath. The heliosphere is asymmetric, pushed inward in the Voyager 2 direction relative to the Voyager 1 direction. The termination shock is a weak, quasi-perpendicular shock that heats the thermal plasma very little. An unexpected finding is that the flow is still supersonic with respect to the thermal ions downstream of the termination shock. Most of the solar wind energy is transferred to the pickup ions or other energetic particles both upstream of and at the termination shock.

The first crossing of the termination shock, by Voyager 1, defined the scale size of the heliosphere; the second crossing, by Voyager 2, reveals the scale of the heliospheric asymmetry. Both termination shock crossings were foreshadowed by observations of streaming energetic particles upstream of the termination shock, in the region named the foreshock in analogy with planetary magnetospheres. Voyager 1, at heliographic latitude 34° N, entered the termination foreshock region at helioradius 85 AU and crossed the termination shock at helioradius 94 AU. Voyager 2, at heliolatitude 26° S, entered the foreshock region at helioradius 75 AU and crossed the termination shock at helioradius 84 AU (see Figs 1, 2 and 3). The spacecraft are

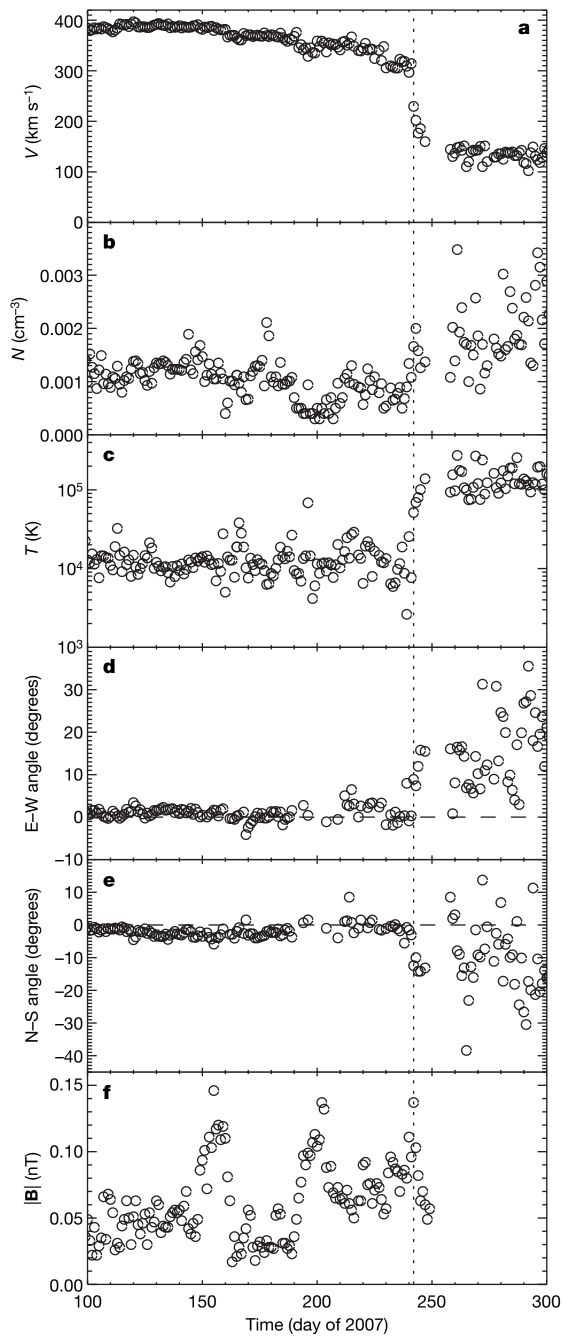


Figure 1 | An overview of data near the termination shock. Daily averages of solar wind speed V (**a**), density N (**b**), temperature T (**c**), east–west flow angle (**d**), north–south flow angle (**e**) and magnetic field magnitude (**f**). Flow angles are in the RTN coordinate system, where R is radially outwards, T is parallel to the plane of the solar equator and positive in the direction of the Sun's rotation, and N completes a right-handed system. The east–west angle is the angle in the R–T plane and the north–south angle is the angle out of the R–T plane. The dashed line shows the termination shock crossing, where the speed decreases by a factor of about two, the density increases by a factor of two, the proton temperature increases to near 100,000 K, and the flow is deflected in the +T and –N directions, consistent with flow away from the nose direction of the heliosphere, that is, the direction toward the local interstellar medium flow. The Voyager plasma experiment measures ions and electrons with energy per electron charge of 10–5,950 eV (ref. 8). Sets of ion and electron spectra were obtained every 192 s; these spectra are fit with convected isotropic maxwellian distributions to determine the plasma velocity, density and temperature.

[1]Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, 37-655, Cambridge, Massachusetts 02139, USA. [2]State Key Laboratory of Space Weather, Center for Space Science and Applied Research, Chinese Academy of Sciences, PO Box 8701, Beijing 100080, China. [3]Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA.

about 45° apart in heliolongitude and are separated by roughly 110 AU. The nearer foreshock boundary, in the Voyager 2 direction, was attributed to the combined effects of a thicker foreshock region in the direction of Voyager 2 and a termination shock that is closer to the sun towards Voyager 2 than it is towards Voyager 1 (refs 7, 8). Interpretations of the interstellar hydrogen and helium flow directions[9,10] and heliospheric radio emissions[11] suggest that the interstellar magnetic field is tilted at 60° from the flow direction of the interstellar medium, which would push the southern portion of the heliosphere inwards relative to the northern portion[7,12], although the magnitude of this asymmetry has been disputed[8,13].

The Voyager 2 data resolve this dispute: Voyager 2 crossed the termination shock 10 AU closer to the Sun than did Voyager 1. However, the termination shock position changes with the solar wind pressure: stronger solar wind pushes the termination shock farther from the Sun. We use the solar wind pressures observed by Voyager 2 before the termination shock crossing and a two-dimensional magnetohydrodynamics model[14] to calculate the motion of the termination shock from the time of the Voyager 1 crossing to that of the Voyager 2 crossing. The termination shock should move in by 2–3 AU owing to solar wind pressure changes; thus, the asymmetry of the termination shock is 7–8 AU, with the shock closer to the Sun in the Voyager 2 direction than it is in the Voyager 1 direction (which could be an east–west or north–south asymmetry, or both).

The first effects of the termination shock on the solar wind speed occurred 0.7 AU upstream of the shock and were very large in comparison with foreshock effects upstream of planetary magnetospheres (see Figs 1 and 4). The solar wind speed decreased from 400 to 300 km s^{-1} in three steps. These steps seemed to be associated with magnetic field enhancements that could be either transient structures or standing waves. Roughly 40% of the bulk energy of the solar wind flow is dissipated before the shock in these speed decreases.

The termination shock was a supercritical quasi-perpendicular shock (see Table 1). The shock curvature is smaller than for a circular shock; this curvature may be due to small-scale oscillations, as the overall shock is expected to be blunt[15]. The shock widths were 300,000 km for TS-2 and 100,000 km for TS-3. These widths are a few times the ion inertial length and are much larger than the electron inertial length and the thermal ion gyroradius.

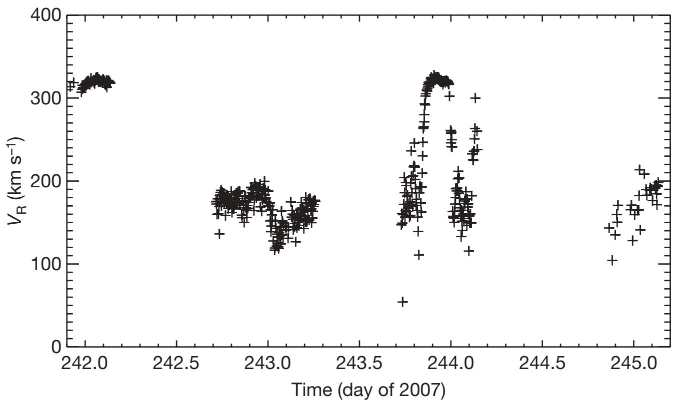


Figure 2 | High-resolution (192 s) solar wind speed (V_R) near the termination shock crossings. At the end of the tracking pass at the beginning of day 242, Voyager 2 was in the solar wind. When tracking resumed at day 242.7, the spacecraft had entered the heliosheath; the first TS crossing (TS-1) occurred in this data gap. Voyager 2 remained in the heliosheath until day 243.85, when the termination shock moved outwards past the spacecraft (the second termination shock crossing, TS-2). Voyager 2 was in the solar wind for about three hours; then the termination shock moved back inward across the spacecraft (TS-3). Another termination shock crossing or partial crossing (TS-4) occurred when the shock again moved outwards at day 244.1. Voyager 2 was in the heliosheath when tracking resumed at the end of day 244 (TS-5 occurred in this gap) and has remained in the heliosheath since that time.

The weakness of the termination shock, with shock compression ratios of 2.4 at TS-2 and 1.6 at TS-3, and the lack of plasma heating at the shock distinguish it from planetary bow shocks, where most of the flow energy is transferred to the thermal plasma (see Fig. 5). The electron temperature at the termination shock also increased by an order of magnitude less (to 3×10^4–4×10^4 K) than at planetary bow shocks. If all the flow energy were transferred to the thermal ions as happens at planetary bow shocks, the heliosheath temperature

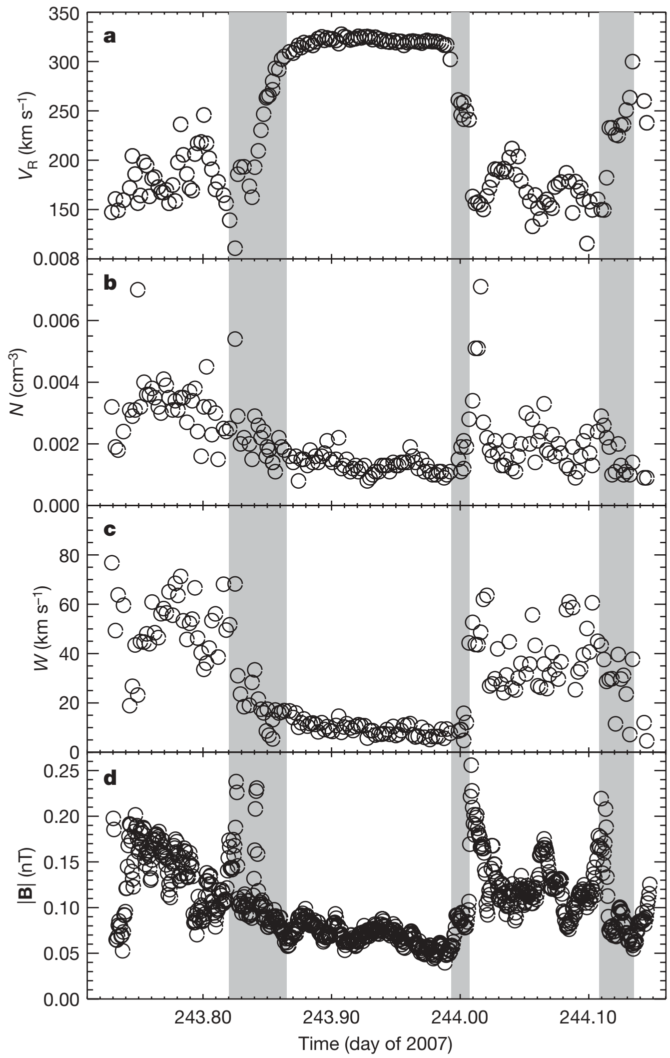


Figure 3 | High-resolution data near the termination shock crossings (shaded regions). Solar wind speed (**a**), density (**b**), thermal speed W (**c**) and magnetic field magnitude (**d**). TS-2, which began at day 243.85, lasted ~1 h. The solar wind speed increased from 150 km s^{-1} to 325 km s^{-1}. The density decreased over a more extended region, from day 283.75 to 283.865, and did not show a sharp jump. The temperature decrease was sharp, but occurred before the speed change and coincided with a peak in the magnetic field strength. The spectra remained roughly maxwellian throughout this period, so plasma changes at the shock occurred on timescales longer than an ion gyroradius. TS-3, at day 244.0, had the structure of a classic supercritical quasi-perpendicular shock[4]. The speed decreased in two steps, first to 250 km s^{-1} in the region called the shock foot, where |**B**| also increased, and then to 150 km s^{-1} at the ramp, the transition to the heliosheath. The density also increased in the shock foot and then again in the heliosheath; the temperature increase occurred at the ramp. About 14 min after the ramp, the heliosheath density increased by a factor of >3 for a few spectra; this feature is not understood. Other very small-scale features are observed in the heliosheath density near the termination shock (for example the single-point peak at day 243.74); these features are real but their source is not known. The crossing at day 244.12 also seems to show a foot–ramp structure; this event may not be a full crossing of the shock but is at least a close encounter.

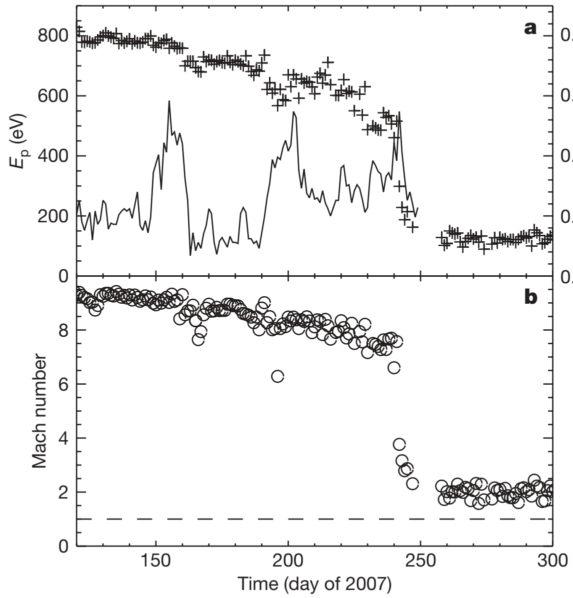


Figure 4 | Most of the solar wind flow energy does not go into the solar wind plasma. **a**, Daily-averaged energy per proton E_p (flow energy plus thermal energy; crosses) near the termination shock, with the magnitude of **B** (continuous trace) superposed. Stepped decreases in energy occurred at days 160, 190 and 230; the first two coincided with increases in the magnetic field strength. **b**, The fast-mode Mach number near the termination shock. The Mach number, the speed divided by the fast-mode wave speed of the thermal plasma, is about nine before the shock and two after the shock. The dashed line shows a Mach number of one. The Mach number was expected to be less than one after the shock.

would be 10^6 K, whereas the observed temperature is 10^5 K. The small temperature increase results in heliosheath flow that is not subsonic with respect to the thermal plasma (see Fig. 4). However, the effect of a shock is to make the flow subsonic (Mach number <1). We explain this discrepancy as follows. The energy per proton decreases by 80% at the termination shock, so the solar wind flow energy is not going into heating the thermal plasma as in planetary bow shocks. This energy is probably transferred to the pickup ions, which are hot protons created when neutral interstellar particles are ionized in the heliosphere. The Voyager 2 plasma data and extrapolations of Voyager 1 energetic particle data to pickup ion energies in the heliosheath are consistent with roughly 80% of the solar wind flow energy being transferred to the pickup ions[16,17]. The wave speeds in the heliosheath would then be determined by the properties of the hot pickup ions, not the thermal plasma, and would be faster than the flow speed, resulting in the heliosheath being subsonic. Particle code models including pickup ions seem consistent with this explanation[18,19].

Table 1 | Termination shock parameters

Parameter	Termination shock motion	
	TS-2: outwards	TS-3: inwards
East–west shock normal angle	$188.0° \pm 4.0°$	$5.8° \pm 10.3°$
Shock speed	94.0 ± 3.4 km s^{-1}	67.9 ± 17.3 km s^{-1}
North–south shock normal angle	$2.0° \pm 6.2°$	$-4.6° \pm 19.2°$
Angle between shock normal and magnetic field	$82.8° \pm 3.9°$	$74.3° \pm 11.2°$
Compression ratio	2.38 ± 0.14	1.58 ± 0.71
Solar wind fast-mode Mach number	4.9 ± 0.1	8.8 ± 1.2
Heliosheath fast-mode Mach number	1.1 ± 0.1	2.8 ± 0.4

The termination shock parameters derived from the Voyager 2 data. We assume that the termination shock is locally planar and described by the fluid magnetohydrodynamics equations. The change in the plasma parameters across the shock must satisfy the Rankine–Hugoniot jump conditions. The most probable orientation of the termination shock is the direction that minimizes the differences of the Rankine–Hugoniot conditions across the shock[20–23]. Once the orientation is determined, the shock speed, the angle between the shock and the magnetic field, the wave speeds, and the Mach numbers are calculated.

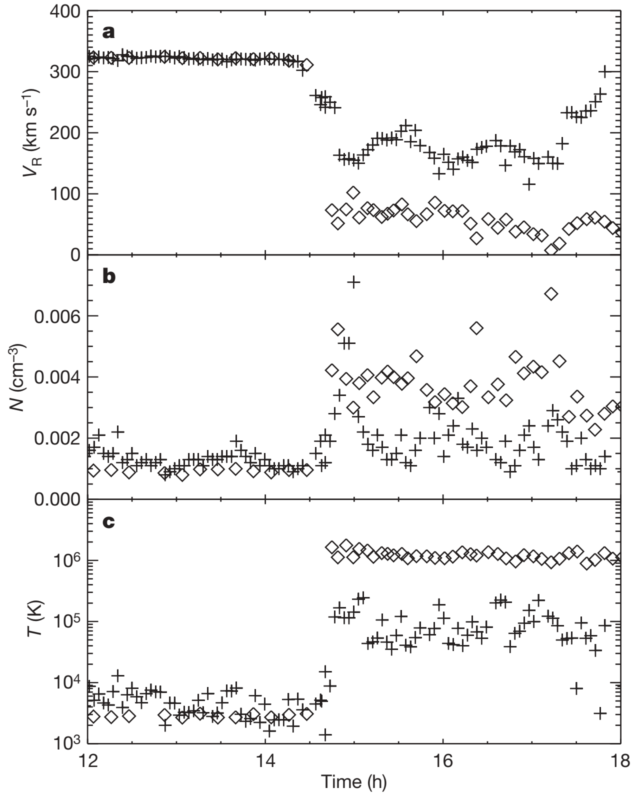


Figure 5 | The termination shock is very different from other shocks observed in the heliosphere. Voyager 2 data measured at TS-2 (crosses) at helioradius 84 AU, in comparison with Voyager 2 data measured at Neptune's inbound bow shock crossing (diamonds) at helioradius 30 AU in August 1989. The solar wind parameters upstream of Neptune are normalized to those upstream of the termination shock; the timescales are identical. The solar wind speed (**a**; Neptune data divided by 1.3) at the bow shock fell by a factor of four but at the termination shock the speed decreased by a factor of only two. The density (**b**; Neptune data divided by five) at the bow shock increased by a factor of four, but at the termination shock by a factor of two. The major difference is in the temperature (**c**; Neptune data divided by two): at the bow shock it increased by a factor of 100, but at the termination shock by a factor of only ten. The differences between these two shocks are probably caused by the greater abundance of pickup ions at the termination shock.

Received 9 March; accepted 15 April 2008.

1. Decker, R. B. *et al.* Voyager 1 in the foreshock, termination shock, and heliosheath. *Science* **309**, 2020–2024 (2005).
2. Burlaga, L. F. *et al.* Crossing the termination shock into the heliosheath: magnetic fields. *Science* **309**, 2027–2029 (2005).
3. Stone, E. C. *et al.* Voyager 1 explores the termination shock region and the heliosheath beyond. *Science* **309**, 2017–2020 (2005).
4. Burlaga, L. F. *et al.* Magnetic fields at the solar wind termination shock. *Nature* doi:10.1038/nature07029 (this issue).
5. Decker, R. B. *et al.* Mediation of the solar wind termination shock by non-thermal ions. *Nature* doi:10.1038/nature07030 (this issue).
6. Gurnett, D. A. & Kurth, W. S. Intense plasma waves at and near the solar wind termination shock. *Nature* doi:10.1038/nature07023 (this issue).
7. Stone, E. C. *et al.* An asymmetric solar wind termination shock. *Nature* doi:10.1038/nature07022 (this issue).
8. Bridge, H. S. *et al.* The plasma experiment on the 1977 Voyager mission. *Space Sci. Rev.* **21**, 259–287 (1977).
9. Opher, M., Stone, E. C. & Liewer, P. C. The effects of a local interstellar magnetic field on Voyager 1 and 2 observations. *Astrophys. J.* **640**, L71–L74 (2006).
10. Pogorelov, N. V., Zank, G. P. & Ogino, T. Three-dimensional features of the outer heliosphere due to coupling between the interstellar and interplanetary magnetic fields. II. The presence of neutral hydrogen atoms. *Astrophys. J.* **644**, 1299–1316 (2006).
11. Lallement, R. *et al.* Deflection of the interstellar neutral hydrogen flow across the heliospheric interface. *Science* **307**, 1447–1449 (2005).
12. Izmodenov, V., Malama, Y. G. & Ruderman, M. Solar cycle influence on the interaction of the solar wind with Local Interstellar Cloud. *Astron. Astrophys.* **429**, 1069–1080 (2005).

13. Gurnett, D. A. & Kurth, W. S. Electron plasma oscillations upstream of the solar wind termination shock. *Science* **309,** 2025–2027 (2005).
14. Opher, M., Stone, E. C. & Gombosi, T. I. The orientation of the local interstellar magnetic field. *Science* **316,** 875–878 (2007).
15. Jokipii, J. R., Giacalone, J. & Kota, J. Transverse streaming anisotropies of charged particles accelerated at the solar wind termination shock. *Astrophys. J.* **611,** L141–L144 (2004).
16. Gloeckler, G., Fisk, L. A. & Lanzerotti, L. J. Acceleration of solar wind and pickup ions by shocks, in *Connecting Sun and Heliosphere (Proc. Solar Wind 11/SOHO 16 Conf.)* (eds Fleck, B., Zurbuchen, T. H. & Lacoste, H.) 107–112 (ESA, Noordwijk, 2005).
17. Zank, G., Pauls, H., Cairns, I. & Webb, G. Interstellar pickup ions and quasi-perpendicular shocks: Implications for the termination shock and interplanetary shocks. *J. Geophys. Res.* **101,** 457–477 (1996).
18. Lipatov, A. S. & Zank, G. P. Pickup ion acceleration at low-β_p perpendicular shocks. *Phys. Rev. Lett.* **82,** 3609–3612 (1999).
19. Giacalone, J. & Burgess, D. Hybrid simulations of the interaction of a current sheet with the termination shock. *Eos* 88 (Fall meeting), abstr. SH12B-05 (2007).
20. Pogorelov, N. V., Stone, E. C., Florinski, V. & Zank, G. P. Termination shock asymmetries as seen by the Voyager spacecraft: The role of the interstellar magnetic field and neutral hydrogen. *Astrophys. J.* **668,** 611–624 (2007).
21. Wang, C., Richardson, J. D. & Paularena, K. I. Predicted Voyager observations of the Bastille Day 2000 coronal mass ejection. *J. Geophys. Res.* **106,** 13007–13013 (2001).
22. Berdichevsky, D. B., Szabo, A., Lepping, R., Viñas, A. & Mariani, F. Interplanetary fast shocks and associated drivers observed through the 23rd solar minimum by Wind over its first 2.5 years. *J. Geophys. Res.* **105,** 27289–27314 (2000).
23. Vinas, A. F. & Scudder, J. D. Fast and optimal solution to the 'Rankine-Hugoniot problem'. *J. Geophys. Res.* **91,** 39–58 (1986).

Acknowledgements The work at MIT is supported by NASA. C.W. is grateful for support from NNSFC. Magnetic field data are shown courtesy of the Voyager magnetometer team (principle investigator N. Ness). We thank G. Gordon, Jr and L. Finck for development of and assistance with the plasma analysis.

Author Contributions J.D.R. analysed the plasma data and wrote the paper. J.C.K. performed the calculations for and write-up of the shock parameters. C.W. calculated the termination shock motion. J.W.B. and A.J.L. assisted with design of the instrument and manuscript preparation.

Author Information Data from the Voyager 2 plasma experiment are available at http://web.mit.edu/space/www/voyager.html. Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.D.R. (jdr@space.mit.edu).

Vol 454 | 3 July 2008 | doi:10.1038/nature07030

LETTERS

Mediation of the solar wind termination shock by non-thermal ions

R. B. Decker[1], S. M. Krimigis[1,2], E. C. Roelof[1], M. E. Hill[1], T. P. Armstrong[3], G. Gloeckler[4], D. C. Hamilton[5] & L. J. Lanzerotti[6,7]

Broad regions on both sides of the solar wind termination shock are populated by high intensities of non-thermal ions and electrons. The pre-shock particles in the solar wind have been measured by the spacecraft Voyager 1 (refs 1–5) and Voyager 2 (refs 3, 6). The post-shock particles in the heliosheath have also been measured by Voyager 1 (refs 3–5). It was not clear, however, what effect these particles might have on the physics of the shock transition until Voyager 2 crossed the shock on 31 August–1 September 2007 (refs 7–9). Unlike Voyager 1, Voyager 2 is making plasma measurements[7]. Data from the plasma[7] and magnetic field[8] instruments on Voyager 2 indicate that non-thermal ion distributions probably have key roles in mediating dynamical processes at the termination shock and in the heliosheath. Here we report that intensities of low-energy ions measured by Voyager 2 produce non-thermal partial ion pressures in the heliosheath that are comparable to (or exceed) both the thermal plasma pressures and the scalar magnetic field pressures. We conclude that these ions are the >0.028 MeV portion of the non-thermal ion distribution that determines the termination shock structure[8] and the acceleration of which extracts a large fraction of bulk-flow kinetic energy from the incident solar wind[7].

We report >0.028 MeV ion and >0.022 MeV electron energy measurements made by the Voyager 2 Low Energy Charged Particle[10] (LECP) instrument in the solar wind, at the termination shock, and in the heliosheath during 2005–2008. We refer to the first crossing of the termination shock by Voyager 2, on day 242 of 2007, as TS-1 (refs 7, 8).

Figure 1b shows that the large energy-dependent variations of low-energy ion intensities measured in the solar wind diminish within 7–8 days of TS-1 occurring. Smaller variations continue in the heliosheath until day 342 of 2007 (2007.94), when a possible transient disturbance reduced intensities of higher energy ions, causing large variations in the spectral index $\gamma(E)$ (Fig. 1c). The low-energy ion energy spectrum measured so far by Voyager 2 in the heliosheath is even flatter than that measured by Voyager 1. For the 123-day period following TS-1 (days 242 to 365 of 2007), Voyager 2 data are well fit by a single power law with index $\gamma = -1.24 \pm 0.05$. By contrast, for the 160-day period following the crossing of the termination shock by Voyager 1 (day 352 of 2004 to day 144 of 2005), that spacecraft's data are well fit by a single power law with index[4] $\gamma = -1.67 \pm 0.03$, which later increased slightly[6] to $\gamma \approx -1.4$ to -1.6.

The salient feature in the LECP measurements of TS-1 is that the non-thermal partial pressure produced by 0.028–3.5 MeV energetic ions in the heliosheath is comparable to (or exceeds) both the thermal plasma pressure[7] (nkT) and the scalar pressure ($B^2/8\pi$) of the magnetic field[8]. The very flat ion spectrum recorded by Voyager 2 enhances the suprathermal partial ion pressure, which is a lower limit on the total particle pressure. The partial pressure >0.028 MeV is compared in Fig. 1a with that measured by Voyager 1. The horizontal dashed line shows the magnetic field pressure calculated using the Voyager 1 mean heliosheath field intensity (at the time of writing, Voyager 2 magnetic field data were unavailable beyond day 245 of 2007). The Voyager 1 partial ion pressure is generally comparable to or exceeds the mean field pressure in the heliosheath. The Voyager 2 partial ion pressure is almost always greater than that measured by Voyager 1 (especially in the pre-shock region). The termination shock and heliosheath regions show high intensities of energetic electrons (Fig. 1d), which were also observed by Voyager 1. The high electron intensities imply that electron impact ionization can have a role in the generation of suprathermal pickup hydrogen that is comparable to (or greater than) that of charge-exchange ionization by the solar wind.

Except for increases due to passage of a merged interaction region and the associated interplanetary shock in March 2006 (ref. 11), intensities of ions <0.2–0.3 MeV in energy remained near background (Fig. 2a, b) until about day 210 of 2007, 32 days before TS-1. This was not the case for Voyager 1, where intensities of 0.04–17.6 MeV ions, averaged over tens of days, increased steadily from mid-2002 to the end of 2004 (refs 4–6). Relatively large intensity increases of 0.14–0.22 MeV ions were recorded by Voyager 1 during the second half of 2002 and during 2004. For ion energies >0.2–0.3 MeV (Fig. 2c–e), upstream intensities are comparable for Voyagers 1 and 2, as Fig. 2e shows; however, over comparable time durations following their respective termination shock crossings, downstream heliosheath intensities observed so far by Voyager 2 are higher than those observed by Voyager 1. Increases in relativistic electrons (Fig. 2f) upstream are less frequent and less intense for Voyager 2 than for Voyager 1, as is true for <0.2–0.3 MeV ions; but downstream they are more intense for Voyager 2 than for Voyager 1, as is true for >0.2–0.3 MeV ions. The variation of the >70 MeV proton count rate (Fig. 2h) indicates modulation of high-energy ion intensities, probably due to solar cycle variations and to passage of transient disturbances, such as the Forbush decrease in March 2006 that was associated with a merged interaction region.

Termination shock ions measured by Voyager 1 in the upstream solar wind between mid-2002 and the end of 2004 were highly anisotropic; they were beamed mainly anti-sunwards along the nearly azimuthal spiral interplanetary magnetic field, that is, in the −T direction[1,3–5]. Those measured by Voyager 2 in the upstream solar wind between the beginning of 2005 and day 242 of 2007 were also anisotropic, but beamed mainly sunwards along the spiral interplanetary magnetic field, that is, in the +T direction, as Fig. 2g shows[3,6]. However, in both the Voyager 2 and Voyager 1 data, anisotropies in

[1]The Johns Hopkins University Applied Physics Laboratory, Laurel, Maryland 20723, USA. [2]Academy of Athens, Soranou Efesiou 4, Athens 11527, Greece. [3]Fundamental Technologies, 2411 Ponderosa, Suite A, Lawrence, Kansas 66046, USA. [4]University of Michigan, 2455 Hayward Street, Ann Arbor, Michigan 48109, USA. [5]University of Maryland, College Park, Maryland 20742, USA. [6]New Jersey Institute of Technology, Newark, New Jersey 07102, USA. [7]Alcatel-Lucent, Murray Hill, New Jersey 07974, USA.

the heliosheath are reduced relative to those in the solar wind, except for the two oppositely directed bursts recorded by Voyager 2 on 2007.9 and 2008.0 in Fig. 2g.

The measurements of high partial pressures of low-energy ions and of high intensities of high-energy electrons made using the Voyager 2 LECP instrument have immediate implications for the nature and processes of the termination shock, the foreshock and the heliosheath. These in turn have ramifications for the global structure of the heliosphere, particle acceleration and propagation processes, as well as for the collisionless shock structure. Voyager 1 entered the termination foreshock proper at a helioradius of 85.2 AU (1 AU is the Sun–Earth distance), and Voyager 2 did so at a helioradius of 75.3 AU, roughly 10 AU nearer the Sun than Voyager 1. First detection by Voyager 2 of termination shock particles around the time that the shock swept over Voyager 1 during its inward movement is qualitatively consistent with three-dimensional heliosphere models which predict an asymmetric termination shock that is nearer the Sun at the position of Voyager 2 because of symmetry-breaking effects of the interstellar magnetic field[12,13]. However, quantitative issues are still under examination[14].

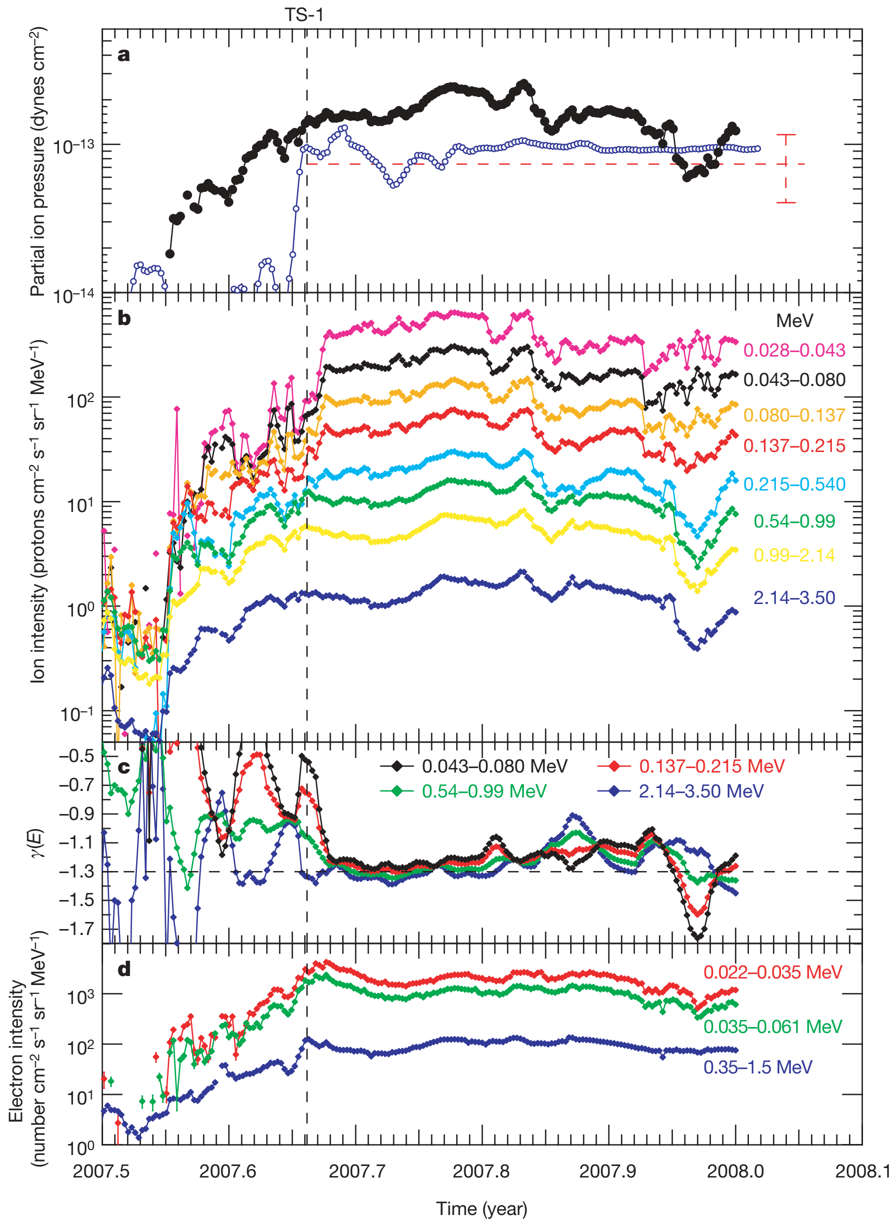
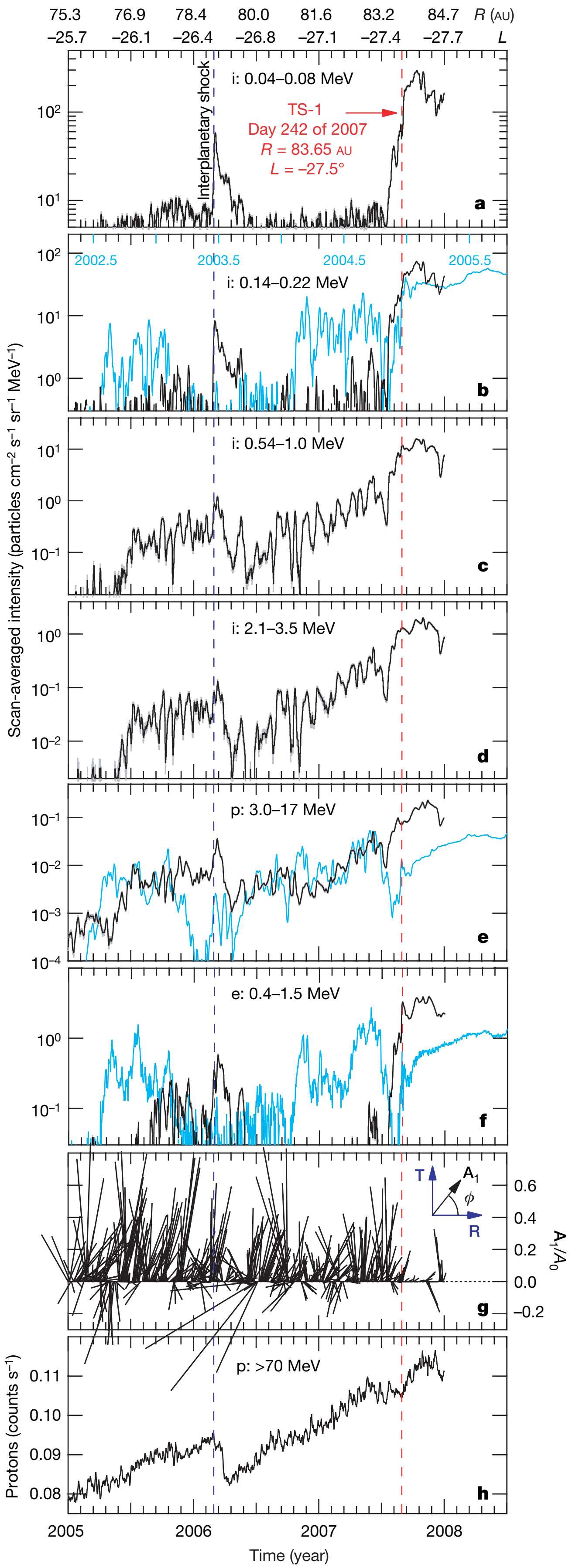


Figure 1 | Low-energy ions and electrons measured by Voyager 2 near the termination shock during 2007. Voyager 2 crossed the termination shock at least five times during days 242 to 244 of 2007 (refs 7–9) at helioradius $R = 83.65$ AU and heliographic latitude $L = -27.5°$. **a**, The black trace shows Voyager 2 0.028–3.5 MeV partial ion (proton) pressures. The blue trace shows Voyager 1 0.040–4.0 MeV partial proton pressures, time-shifted so that the termination shock crossing of Voyager 1 on day 351 of 2004 coincides with TS-1. Both traces exceed the magnetic field pressure $P_{B_1} = B_1^2/8\pi$ (dashed red line), calculated using the mean heliosheath field intensity $B_1 = 0.123 \pm 0.035$ nT (± 1 s.d.) measured by Voyager 1 between days 1 and 110 of 2005 (ref. 20). The upper and lower bounds of the vertical error bar are P_{B_1} evaluated respectively at $B_1 = 0.158$ and 0.088 nT. The Voyager 1 and 2 ion pressures were calculated using intensities of ions arriving from the sunwards and anti-sunwards directions, to reduce contributions from field-aligned beams that arrive mainly from the azimuthal direction. **b**, Intensities in the eight Voyager 2 ion channels calculated using proton energy passbands and efficiencies. **c**, Differential spectral index $\gamma(E)$ evaluated at the logarithmic means of the energy passband for the channels indicated[6]. **d**, 0.022–1.5 MeV electron intensities; ordinate of blue trace is multiplied by 30. Electron intensities peak at or near the termination shock, are nearly isotropic in the solar wind and heliosheath, and at ~0.03 MeV and ~0.7 MeV are respectively higher by a factor of ~4 or lower by a factor of ~3 than intensities of ions at the same energy. All data shown are one-day averaged; data in **c** are also five-point smoothed.



Numerical models had predicted that Voyager 2 should observe termination shock ion anisotropies directed towards the Sun (the +T direction), in the opposite direction to those observed by Voyager 1 (ref. 12), which is consistent with the data. Currently, the models do not explain why the intensities of termination shock ions observed by Voyager 2 in the solar wind are deficient at energies <0.2–0.3 MeV until only ~30 days before TS-1. One possibility is that shock normal vectors in regions of the termination shock to which Voyager 2 was magnetically connected were, on average, more nearly perpendicular to the field than were those to which Voyager 1 was connected, thus hindering the upstream escape of slower, lower energy ions from the termination shock, and reducing their intensities at Voyager 2 (ref. 15). Another possibility, suggested by recent modelling of termination shock motion[16], is that during the ~2.5 yr periods when upstream termination shock particles were observed at the spacecraft, Voyager 1 remained relatively close to the termination shock in comparison with Voyager 2. However, the intensities of 0.35–1.5 MeV electrons (speeds ~160 AU d^{-1}) in the solar wind measured by Voyager 2 also are lower than those measured by Voyager 1 (Fig. 2f). This might suggest that the electron source region is more remote from the spacecraft than is that of the termination shock ions.

Suprathermal ions in the heliosheath probably originate as solar wind pickup ions that undergo acceleration or heating, or both, at the termination shock and in the heliosheath[17–19]. In the simplest application of diffusive shock acceleration theory, a spectral index of $\gamma \approx -1.25$ for 0.028–3.5 MeV heliosheath ions implies a density compression ratio of ~3 across the termination shock, which is inconsistent with the value of ~2 deduced from Voyager 2 plasma and field data[7,8]. This spectral index is also inconsistent with the value of $\gamma = -1.5$ predicted when the suprathermal tail of the solar wind pickup protons is heated across the termination shock[19]. We note that these current theories and models are applicable under conditions that are steady in time and uniform in space. Voyager 2 is still evidently in the region of the heliosheath near the termination shock, and it is not clear from the Voyager 2 data that either of the conditions is met. More importantly, we stress that the estimated partial ion pressure derived from the LECP data in Fig. 1a considerably exceeds the magnetic pressure and is consistent with other Voyager 2 plasma[7] and magnetic field[8] data which indicate that energized

Figure 2 | Energetic ions and electrons measured by Voyager 2 during 2005–2007. Expanded view showing Voyager 2 LECP data during 2005–2007, when Voyager 2 moved from $R = 75.3$ AU, $L = -25.7°$ to $R = 84.7$ AU, $L = -27.7°$. Instruments on Voyager 1 began measuring high intensities of termination shock particles upstream of the shock in mid 2002 at $R = 85.2$ AU, $L = 33.8°$ (refs 1, 2). Voyager 1 crossed the termination shock at $R = 94.0$ AU on day 351 of 2004 (refs 3–5, 20, 21) as the shock moved radially inwards as a result of decreasing solar wind ram pressure[22]. The asymmetry of the termination shock allowed Voyager 2, then in the solar wind at only $R = 75.3$ AU, $L = -25.6°$, to enter the termination foreshock, where the Voyager 2 instruments began measuring termination shock particles[3,6]. **a**–**f**, Black traces are Voyager 2 LECP scan-averaged intensities at indicated particle energies (i, ion; p, proton; e, electron); grey error bars show 2s.d. Blue traces (**b**, **e**, **f**) are Voyager 1 LECP intensities shifted in time by +2.705 yr to align the termination shock crossings of Voyager 1 (day 351 of 2004) and Voyager 2 (day 242 of 2007); Voyager 1 data shown cover the period between day 109 of 2002 and day 291 of 2005. The time (in years) for the Voyager 1 data are indicated along the top axis of **b**. **g**, First-order anisotropy vector $\mathbf{A}_1/A_0$ derived from a least-squares Fourier fit in the R–T plane (unit vector **R** points radially outward from Sun, unit vector **T** points in the direction of normal planetary motion) of measured intensities in seven of eight 45° sectors for proton intensity in **e**[4,5]. The $\mathbf{A}_1/A_0$ 'whiskers' point in the direction in which the particles are moving, and ϕ is the azimuthal angle in the R–T plane[4]. **h**, Count rate of >70 MeV protons. The rapid decrease followed by partial recovery at the end of 2007 suggests that similar intensity decreases seen in all LECP channels are caused by passage of a transient disturbance through the heliosheath. Voyager 2 helioradius R and heliographic latitude L are indicated at half-year intervals along the top of **a**. All data shown are one-day averaged and five-point smoothed.

pickup ions (as measured by the LECP instrument) have a major dynamical role in mediating the termination shock structure and its variability, thereby significantly affecting dynamical processes in the heliosheath plasma[7,8].

Received 21 February; accepted 15 April 2008.

1. Krimigis, S. M. *et al.* Voyager 1 exited the solar wind at a distance of ~85 AU from the Sun. *Nature* **426,** 45–48 (2003).
2. McDonald, F. B. *et al.* Enhancements of energetic particles near the heliospheric termination shock. *Nature* **426,** 48–51 (2003).
3. Stone, E. C. *et al.* Voyager 1 explores the termination shock region and the heliosheath beyond. *Science* **309,** 2012–2020 (2005).
4. Decker, R. B. *et al.* Voyager 1 in the foreshock, termination shock, and heliosheath. *Science* **309,** 2020–2024 (2005).
5. Krimigis, S. M. *et al.* in *Connecting Sun and Heliosphere (Proc. Solar Wind 11/SOHO 16 Conf.)* (eds Fleck, B., Zurbuchen, T. H. & Lacoste, H.) 21–28 (ESA, Noordwijk, 2005).
6. Decker, R. B. *et al.* in *Physics of the Inner Heliosheath: Voyager Observations, Theory, and Future Prospects (Proc. 5th IGPP Internat. Astrophys. Conf.)* (eds Heerikhuisen, J., Florinski, V., Zank, G. P. & Pogorelov, N. V.) 73–78 (AIP, Melville, New York, 2006).
7. Richardson, J. D. *et al.* Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock. *Nature* doi:10.1038/nature07024 (this issue).
8. Burlaga, L. F. *et al.* Magnetic fields at the solar wind termination shock. *Nature* doi:10.1038/nature07029 (this issue).
9. Gurnett, D. A. & Kurth, W. S. Intense plasma waves at and near the solar wind termination shock. *Nature* doi:10.1038/nature07023 (this issue).
10. Krimigis, S. M. *et al.* The Low Energy Charged Particle (LECP) experiment on the Voyager spacecraft. *Space Sci. Rev.* **21,** 329–354 (1977).
11. Richardson, J. D. *et al.* Source and consequences of a large shock near 79 AU. *Geophys. Res. Lett.* **33,** L23107 (2006).
12. Opher, M., Stone, E. C. & Liewer, P. C. The effects of a local interstellar magnetic field on Voyager 1 and 2 observations. *Astrophys. J.* **640,** L71–L74 (2006).
13. Pogorelov, N. V. in *Physics of the Inner Heliosheath: Voyager Observations, Theory, and Future Prospects (Proc. 5th IGPP Internat. Astrophys. Conf.)* (eds Heerikhuisen, J., Florinski, V., Zank, G. P. & Pogorelov, N. V.) 3–13 (AIP, Melville, New York, 2006).
14. Pogorelov, N. V., Stone, E. C., Florinski, V. & Zank, G. P. Termination shock asymmetries as seen by the Voyager spacecraft: The role of the interstellar magnetic field and neutral hydrogen. *Astrophys. J.* **668,** 611–624 (2007).
15. Jokipii, J. R., Giacalone, J. & Decker, R. B. Energy spectra of energetic particles upstream of the termination shock. *Eos* **88** (Fall meeting), abstr. SH11A–07 (2007).
16. Washimi, H. *et al.* A forecast of the heliospheric termination-shock position by three-dimensional MHD simulations. *Astrophys. J.* **670,** L139–L142 (2007).
17. le Roux, J. A., Webb, G. M., Florinski, V. & Zank, G. P. A focused transport approach to pickup ion shock acceleration: Implications for the termination shock. *Astrophys. J.* **662,** 350–371 (2007).
18. Giacalone, J. & Jokipii, J. R. Energetic particles around the termination shock: Numerical simulations for a blunt shock with cross-field diffusion. *Astrophys. J.* **649,** L137–L140 (2006).
19. Fisk, L. A., Gloeckler, G. & Zurbuchen, T. H. Acceleration of low-energy ions at the termination shock of the solar wind. *Astrophys. J.* **644,** 631–637 (2006).
20. Burlaga, L. F. *et al.* Crossing the termination shock into the heliosheath: magnetic fields. *Science* **309,** 2027–2029 (2005).
21. Gurnett, D. A. & Kurth, W. S. Electron plasma oscillations upstream of the solar wind termination shock. *Science* **309,** 2025–2027 (2005).
22. Richardson, J. D. *et al.* Relation between the solar wind dynamic pressure at Voyager 2 and the energetic particle events at Voyager 1. *J. Geophys. Res.* **110,** A09106 (2005).

Acknowledgements Work at the Johns Hopkins University Applied Physics Laboratory was supported by the Voyager Interstellar Mission under NASA grant NNX07AB02G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.B.D. (robert.decker@jhuapl.edu).

nature

An asymmetric solar wind termination shock

Edward C. Stone[1], Alan C. Cummings[1], Frank B. McDonald[2], Bryant C. Heikkila[3], Nand Lal[3] & William R. Webber[4]

Voyager 2 crossed the solar wind termination shock at 83.7 AU in the southern hemisphere, ~10 AU closer to the Sun than found by Voyager 1 in the north[1–4]. This asymmetry could indicate an asymmetric pressure from an interstellar magnetic field[5,6], from transient-induced shock motion[7], or from the solar wind dynamic pressure. Here we report that the intensity of 4–5 MeV protons accelerated by the shock near Voyager 2 was three times that observed concurrently by Voyager 1, indicating differences in the shock at the two locations. (Companion papers report on the plasma[8], magnetic field[9], plasma-wave[10] and lower energy particle[11] observations at the shock.) Voyager 2 did not find the source of anomalous cosmic rays at the shock, suggesting that the source is elsewhere on the shock[12–14] or in the heliosheath[15–19]. The small intensity gradient of Galactic cosmic ray helium indicates that either the gradient is further out in the heliosheath[20] or the local interstellar Galactic cosmic ray intensity is lower than expected[21].

Low energy ions accelerated at the termination shock are observed upstream of the shock and in the heliosheath (Fig. 1). Voyager 2

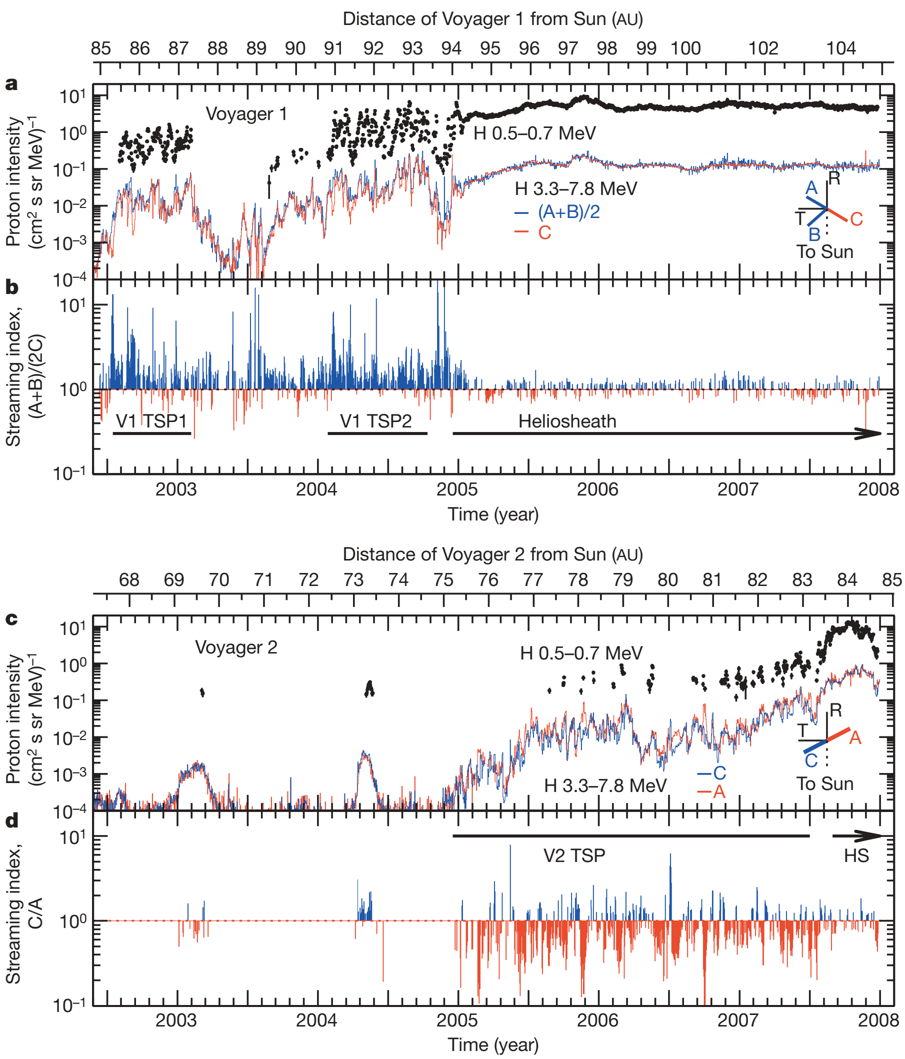


Figure 1 | Daily-averaged intensities and streaming of energetic termination shock particles that are accelerated at nearby regions of the shock. Voyager 1 and Voyager 2 crossed the shock and entered the heliosheath on 2004.96 (16 December 2004) at heliographic coordinates of (34.3°, 173°) and on 2007.66 (30 August 2007) at (−27.5°, 216°), respectively. Insets, telescope (A, B and C) viewing directions projected into the R–T plane, where −R is towards the Sun and T is azimuthal. Error bars on black filled circles, ±1s.d. **a**, The proton intensities (H) at 3.3–7.8 MeV observed by Voyager 1 particle telescopes (A+B)/2 (blue trace) and by C (red trace) are highly variable upstream of the shock owing to variations in the connectivity along the spiral field line[28,29]. The energetic ions are convected into the heliosheath, resulting in reduced variations. Similar properties are apparent in the intensity of 0.5–0.7 MeV protons observed by telescope A (black filled circles) and shown when the background correction was <60%. V1 TSP1 and V1 TSP2, two episodes of termination shock particles observed by Voyager 1. **b**, The streaming index (A+B)/(2C) for 3.3–7.8 MeV protons shows that upstream the ions at Voyager 1 were strongly beamed in the –T direction, with intensities in the oppositely directed detectors differing by up to a factor of 10. The intensities are more nearly isotropic in the heliosheath. Blue indicates that the average intensity in telescopes A and B exceeds that in C, indicating flow in the –T direction; red indicates the opposite. **c**, Same as **a** for Voyager 2 except that only telescopes A and C are used in determining the directional intensities of 3.3–7.8 MeV protons. **d**, Same as **b** for Voyager 2 except that only telescopes A and C are used. The upstream beaming was mainly in the +T direction, opposite to that observed by Voyager 1 and consistent with the predicted east–west shock asymmetry resulting from a local interstellar magnetic field[5,6,30]. Voyager 2 began observing upstream energetic ions at 75 AU from the Sun[1], 10 AU closer than did Voyager 1, leading to predictions that the shock would be closer in the southern hemisphere than in the north, but with significant differences in the predicted asymmetry[5–7]. HS, heliosheath.

[1]California Institute of Technology, Pasadena, California 91125, USA. [2]Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA. [3]NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [4]Department of Physics and Astronomy, New Mexico State University, Las Cruces, New Mexico 88003, USA.

began observing upstream ions at 75 AU, leading to model predictions, differing in detail, that the shock would be found closer to the Sun by Voyager 2 than by Voyager 1 (refs 5–7). The observed shock location[8–11], low heliosheath plasma temperature[8], and high energetic ion pressure[11] will lead to improvements in the models.

There has been little variation in the intensity of termination shock particles (TSPs) observed by Voyager 1 in the heliosheath since mid-2005. However, the intensity of helium nuclei with an energy of 1–1.5 MeV per nucleon and 4–5 MeV protons at Voyager 2 just after it crossed the termination shock was three times larger than Voyager 1 observed concurrently (Fig. 2), suggesting that shock conditions affecting acceleration vary with time or shock region. Given the stability of the TSP intensity at Voyager 1, it will be important to determine the evolution of the spectrum at Voyager 2.

Among the surprises from the Voyager 1 termination shock crossing was that the intensity of anomalous cosmic rays (ACRs) did not peak at the shock as predicted[1]. Models have been proposed in which transients modify acceleration by the shock[22–24], the source of higher energy ACRs is the flank or tail region of the shock[12–14], or acceleration occurs as the particles diffuse outward in the heliosheath[15–19].

The Voyager 1 and Voyager 2 ACR proton and helium spectra just after the latter's shock crossing show distinct differences from those in late 2004 just after Voyager 1 crossed the shock (Fig. 2). The similarity of the changes at Voyager 1 in the heliosheath and Voyager 2 upstream (Figs 2 and 3) between the two spacecrafts' shock crossings suggests a common temporal change due to decreasing solar modulation. As a result, the Voyager 2 intensity of 12–22 MeV per nucleon He just after the shock was 8.4 ± 1.8 times the intensity at Voyager 1 just after its crossing. It was, however, a factor of 2.5 ± 0.4 smaller than the concurrent Voyager 1 intensity in the heliosheath, indicating that Voyager 2 did not observe the expected ACR source spectrum near the shock.

The ratio of the Voyager 1 and Voyager 2 intensities is a measure of the gradient between the spacecraft. The energy dependence of the proton and helium ratios (Fig. 2c, d) is essentially the same, but with the energy scales differing by a factor of four. Such scaling is expected from diffusive processes involving ions with velocity v, mean free path λ, and diffusion coefficient $\kappa = v\lambda/3$. If the dependence of λ on the ion momentum per charge is $\lambda \propto R^{\gamma}$, where the rigidity $R = Mv/(Qe)$, then two species with masses M of M_1 and M_2 and charges Q of Q_1 and Q_2 will have the same κ if their energies per nucleon are scaled as $E_1/E_2 = ((M_1/Q_1)/(M_2/Q_2))^{2\gamma/(\gamma+1)}$. For singly ionized ACRs, an energy scaling of 4 to 5 (ref. 25) indicates that $\lambda \approx R$ to $R^{1.4}$ for $0.2 < R < 1.5$ GV. Thus, the scaling in Fig. 2 probably reflects a diffusive spatial gradient in the heliosheath and a remote ACR source at the flank or tail region of the shock or further out in the heliosheath.

The ACR gradients in the heliosheath are also apparent in Fig. 3a. The absence of a gradient between Voyager 1 and Voyager 2 for 61–73 MeV per nucleon ACRs since mid-2005 indicates that the mean free path for ions with rigidities $R > 1.4$ GV is sufficiently large that the intensity is uniform in the nose region of the heliosheath and is probably the ACR source intensity. The constancy of the intensity indicates a steady, high energy ACR source since at least mid-2005.

Shorter diffusive mean free paths for lower rigidities result in intensity gradients for lower energy ACRs. The Voyager 2/Voyager 1 He intensity ratio of 0.4 ± 0.1 at 12–22 MeV per nucleon (Fig. 2d) corresponds to a radial gradient of 4.5% per AU in the heliosheath. The radial gradient would be smaller if some of the difference between Voyager 2 and Voyager 1 was due to longitudinal or latitudinal gradients (see, for example, ref. 13).

A gradient of 4.5% per AU is reasonably consistent with the intensity increase of 12–22 MeV per nucleon He at Voyager 1 between 97.0 and 102.4 AU (2005.8 and 2007.3; Fig. 3a), but not with the factor of 10 increase that occurred in the 10 months immediately following the shock crossing. This suggests that the latter increase was temporal, not spatial as has been assumed in fitting stochastic acceleration models to the Voyager 1 observations. Another suggestion to explain

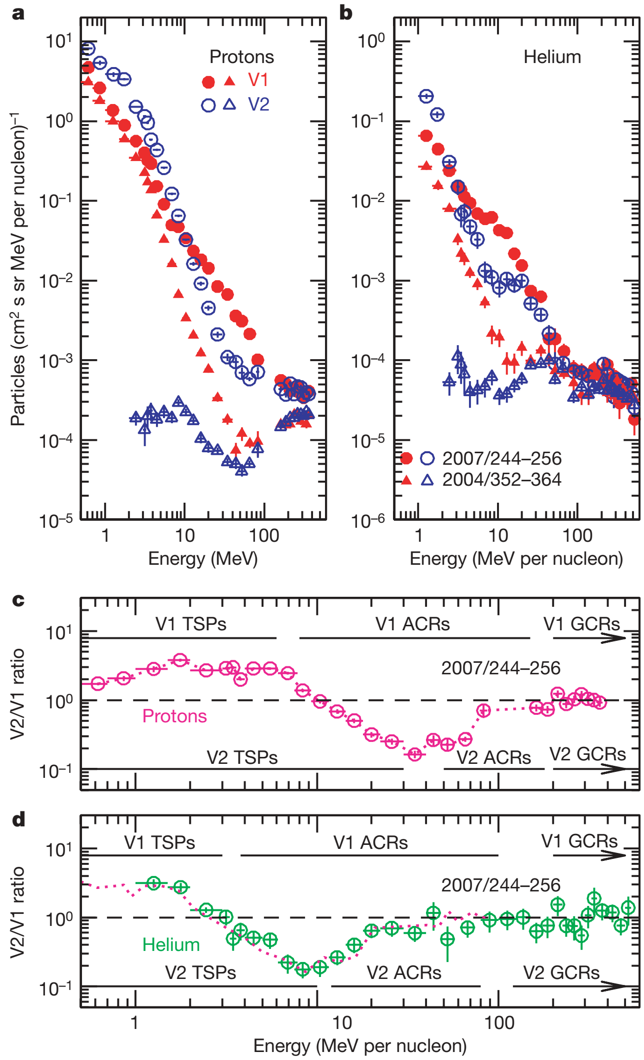


Figure 2 | Comparison of the energy spectra of protons and helium nuclei in the heliosheath near the times of the Voyager 1 and Voyager 2 shock crossings. During the 2004 period, Voyagers 1 and 2 were at 94.1 and 75.2 AU, respectively, and during the 2007 period, the former was at 103.8 AU and 34.3° N heliographic latitude and the latter was at 83.7 AU and 27.5° S. Error bars, ±1s.d. **a**, Proton spectra. Three components are apparent in 2007: TSPs convected from nearby shock regions (<6 MeV), anomalous cosmic rays (ACRs) that are modulated in the heliosheath (8–150 MeV), and Galactic cosmic rays (GCRs) (>200 MeV). The ACR and GCR intensities increased from 2004 to 2007 as solar modulation decreased with declining solar activity. In 2007, the Voyager 2 (V2) TSP intensity at 4–5 MeV was three times that at Voyager 1 (V1) and the spectrum from 0.5 to 3 MeV was $\sim E^{-1.1\pm0.1}$, harder than the Voyager 1 spectrum. The TSP spectra have breaks at ~3 MeV, with the Voyager 2 spectrum falling as $E^{-3.1\pm0.1}$ from 5–30 MeV. Triangles, spectra during period immediately following Voyager 1 shock crossing; circles, spectra during period immediately following Voyager 2 crossing. **b**, Helium spectra with time periods indicated (as year/days of year). Three helium components are apparent, although the energy intervals differ. In the 2004 period, the ACR component was observable down to ~10 MeV per nucleon. By the 2007 period, the intensity of 12–22 MeV per nucleon ACR He at Voyager 1 increased by a factor of 21 ± 4 as the spectrum approached the expected power law source spectrum. The intensity at Voyager 2 increased by a factor of 20 ± 3, suggesting a common change in solar modulation. Filled symbols, Voyager 1; open symbols, Voyager 2. **c**, Ratio of the proton intensities in **a** for 2007. The horizontal lines mark energies dominated by TSPs, ACRs and GCRs. The dotted line is drawn through the Voyager data points for reference. **d**, Ratio of the helium intensities in 2007. The dotted line is the proton intensity ratio from **c** shifted by a factor of 4 in energy per nucleon, showing that the proton and helium ratios are essentially the same. The dashed lines in **c** and **d** correspond to equal intensities at Voyager 1 and 2.

the Voyager 1 observations was the effect of a large transient on shock acceleration[22]. However, from the 12–22 MeV per nucleon He intensity profile at Voyager 2 in Fig. 3a, it does not appear that there was a large enough transient effect at the time of the Voyager 2 shock crossing to support that suggestion.

The heliosheath is expected to impede the diffusion of low energy Galactic cosmic rays (GCRs) into the heliosphere, resulting in a positive radial gradient. The estimated[21] local interstellar intensity of GCR He with 150–380 MeV per nucleon is twice the intensity observed by Voyager 1 at 104 AU (Fig. 3a), requiring a gradient of 2.2% per AU if it were uniform between Voyager 1 and a heliopause at 135 AU. However, the Voyager 1 and Voyager 2 intensities correspond to a radial gradient of $-0.2 \pm 0.2\%$ per AU, indicating there is either a larger gradient beyond 105 AU (ref. 20) or the local interstellar intensity is lower than current estimates.

The cosmic ray instrument also measures the intensity of electrons with energies from 2.5 to 160 MeV (Fig. 3b). Because of their low rigidities, GCR electrons with these energies are strongly modulated and their intensities should be larger in the outer heliosphere. They are also strongly affected by transients. However, the temporal effects associated with the two large transient magnetic field increases just upstream of the shock around 2007.42 and 2007.55 differed from those of merged interaction regions previously observed by Voyager 1 and Voyager 2 (ref. 26). These recent Voyager 2 transients did not produce concurrent transient intensity increases of 10 MeV electrons and 2–50 MeV H ions, nor was there an increase in the solar wind dynamic pressure[8], suggesting that they may be of a different character and possibly related to the nearby presence of the termination shock.

The Voyager 2 electron intensity at 2.5–5.2 MeV continued to increase after the shock crossing, reaching that concurrently observed by Voyager 1 at 104 AU. However, the Voyager 2 intensities at 6–14 MeV and at 26–45 MeV were less than at Voyager 1, indicating that the energy spectrum observed by the former spacecraft near the shock in the heliosheath is much steeper than observed concurrently by Voyager 1 much further from the shock. This suggests that other processes, such as re-acceleration at the shock[27], may be occurring.

With both Voyager 1 and Voyager 2 now in the heliosheath, it will be possible to determine the intensity gradients of ACRs and GCRs

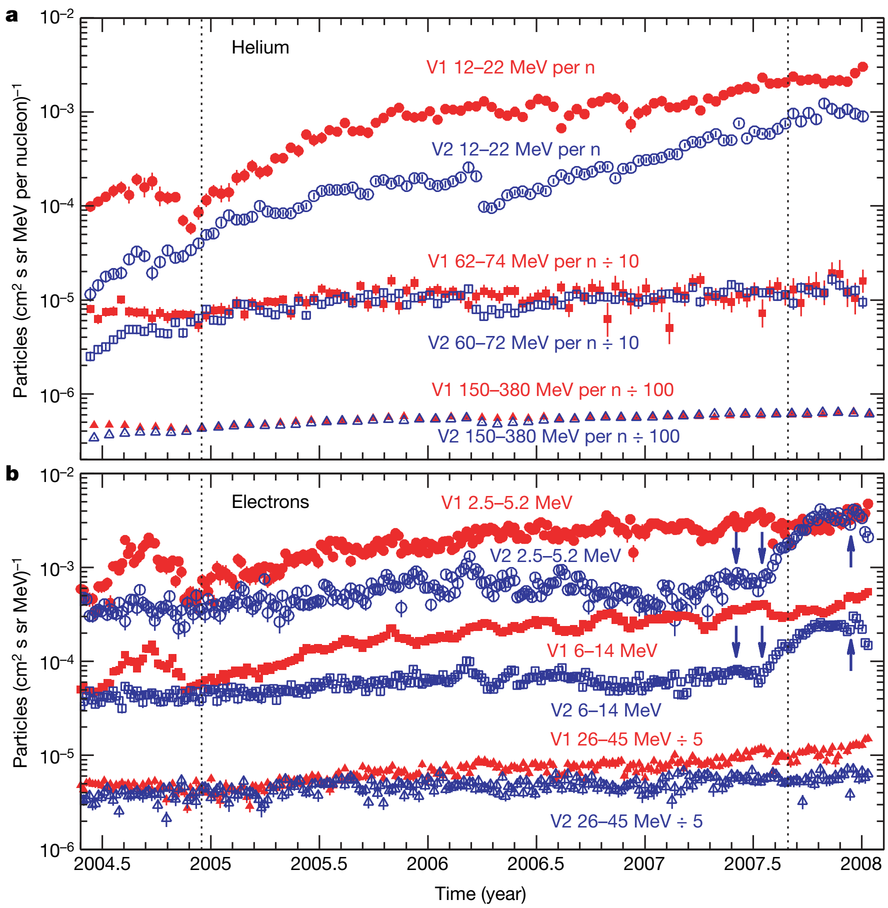


Figure 3 | Temporal changes in the intensities of helium nuclei and electrons. The vertical dotted lines mark times of the Voyager 1 shock crossing (2004.96) and the Voyager 2 crossing (2007.66), after which both spacecraft have been in the heliosheath. Error bars, ±1s.d. **a**, Intensities of helium nuclei in three energy bands. ACRs dominate in the two lowest energy bands plotted, and GCRs dominate in the highest energy band. The Voyager 1 and Voyager 2 intensities at ~61–73 MeV per nucleon are essentially the same, not only in 2008 when both are in the heliosheath, but for the past three years when Voyager 2 was upstream of the shock. In addition, the ~61–73 MeV per nucleon intensity has been nearly constant since mid-2005, indicating a steady ACR source at high energies and no observable gradient between the two spacecraft. There is an ACR gradient in the heliosheath at 12–22 MeV per nucleon as indicated by the difference in the intensities at Voyager 1 and Voyager 2. The nearly identical intensities of GCR helium nuclei with 150–380 MeV per nucleon indicate that the radial gradient in the heliosheath is $-0.2 \pm 0.2\%$ per AU between Voyager 2 at 84 AU and Voyager 1 at 104 AU. **b**, Intensities of electrons in three energy bands. Arrows indicate three transient events that affected the electron intensities. The intensities at Voyager 2 began increasing rapidly just upstream of the shock following the passage of two large transient magnetic field increases[9] around 2007.42 and 2007.55. About 50 days later, the Voyager 2 intensity equalled that at Voyager 1 at 2.5–5.2 MeV, but not at higher energies. However, there was a transient in the heliosheath[8] starting at 2007.95, with an associated transient intensity increase at energies up to 16 MeV, followed by a sharp decrease. This is similar to the passage of merged interaction regions previously observed in the supersonic solar wind.

during the next year, when there will be a minimum level of solar modulation, and to observe the evolution of the gradients as the level of modulation increases with the onset of the new cycle of solar activity.

Received 21 February; accepted 15 April 2008.

1. Stone, E. C. *et al.* Voyager 1 explores the termination shock region and the heliosheath beyond. *Science* **309,** 2017–2020 (2005).
2. Burlaga, L. F. *et al.* Crossing the termination shock into the heliosheath: Magnetic fields. *Science* **309,** 2027–2029 (2005).
3. Decker, R. B. *et al.* Voyager 1 in the foreshock, termination shock, and heliosheath. *Science* **309,** 2020–2024 (2005).
4. Gurnett, D. A. & Kurth, W. S. Electron plasma oscillations upstream of the solar wind termination shock. *Science* **309,** 2025–2027 (2005).
5. Opher, M., Stone, E. C. & Liewer, P. C. The effects of a local interstellar magnetic field on Voyager 1 and 2 observations. *Astrophys. J.* **640,** L71–L74 (2006).
6. Pogorelov, N. V., Stone, E. C., Florinski, V. & Zank, G. P. Termination shock asymmetries as seen by the Voyager spacecraft: The role of the interstellar magnetic field and neutral hydrogen. *Astrophys. J.* **668,** 611–624 (2007).
7. Washimi, H., Zank, G. P., Hu, Q., Tanaka, T. & Munakata, K. A forecast of the heliospheric termination-shock position by three-dimensional MHD simulations. *Astrophys. J.* **670,** L139–L142 (2007).
8. Richardson, J. D., Kasper, J. C., Wang, C., Belcher, J. W. & Lazarus, A. J. Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock. *Nature* doi:10.1038/nature07024 (this issue).
9. Burlaga, L. F. *et al.* Magnetic fields at the solar wind termination shock. *Nature* doi:10.1038/nature07029 (this issue).
10. Gurnett, D. A. & Kurth, W. S. Intense plasma waves at and near the solar wind termination shock. *Nature* doi:10.1038/nature07023 (this issue).
11. Decker, R. B. *et al.* Mediation of the solar wind termination shock by non-thermal ions. *Nature* doi:10.1038/nature07030 (this issue).
12. McComas, D. J. & Schwadron, N. A. An explanation of the Voyager paradox: Particle acceleration at a blunt termination shock. *Geophys. Res. Lett.* **33,** L04102 (2006).
13. Kóta, J. Anomalous cosmic rays at a blunt termination shock. *Proc. 30th Int. Cosmic Ray Conf.* (in the press).
14. Schwadron, N. A., Lee, M. A. & McComas, D. J. Diffusive acceleration at the blunt termination shock. *Astrophys. J.* **675,** 1584–1600 (2008).
15. Ferreira, S. E. S., Potgieter, M. S. & Scherer, K. Transport and acceleration of anomalous cosmic rays in the inner heliosheath. *J. Geophys. Res.* **112,** A11101, doi:10.1029/2007JA012477 (2007).
16. Fisk, L. A., Gloeckler, G. & Zurbuchen, T. H. Acceleration of low-energy ions at the termination shock of the solar wind. *Astrophys. J.* **644,** 631–637 (2006).
17. Langner, U. W. & Potgieter, M. S. Possible explanations of anomalous spectra observed with Voyager 1 crossing the solar wind termination shock. *AIP Conf. Proc.* **858,** 233–238 (2006).
18. Moraal, H. *et al.* Cosmic ray energy changes at the termination shock and in the heliosheath. *AIP Conf. Proc.* **858,** 219–225 (2006).
19. Zhang, M. Acceleration of galactic and anomalous cosmic rays in the heliosheath. *AIP Conf. Proc.* **858,** 226–232 (2006).
20. Florinski, V. *et al.* Galactic cosmic ray transport in the global heliosphere. *J. Geophys. Res.* **108,** A1228, doi:10.1029/2002JA009695 (2003).
21. Webber, W. R. & Lockwood, J. A. Voyager and Pioneer spacecraft measurements of cosmic ray intensities in the outer heliosphere: Toward a new paradigm for understanding the global solar modulation process: 1. Minimum solar modulation (1987 and 1997). *J. Geophys. Res.* **106,** 29323–29332 (2001).
22. Florinski, V. & Zank, G. P. Particle acceleration at a dynamic termination shock. *Geophys. Res. Lett.* **33,** L15110 (2006).
23. Jokipii, J. R. Energetic particles near and beyond the solar-wind termination. *AIP Conf. Proc.* **858,** 143–152 (2006).
24. Kóta, J. & Jokipii, J. R. Implications of the Voyager-1 particle spectra for acceleration at the termination shock. *AIP Conf. Proc.* **858,** 171–176 (2006).
25. Cummings, A. C. & Stone, E. C. Possible role of transients on the energy spectra of energetic particles at the solar wind termination shock. *Proc. 30th Int. Cosmic Ray Conf.* (in the press).
26. McDonald, F. B. *et al.* Observations of energetic ions and electrons in the distant heliosphere: 2001 - 2005.0. *AIP Conf. Proc.* **781,** 261–266 (2005).
27. Ferreira, S. E. S. & Potgieter, M. S. The modulation of 4- to 16-MeV electrons in the outer heliosphere: Implications of different local interstellar spectra. *J. Geophys. Res.* **107,** A1221, doi:10.1029/2001JA000226 (2002).
28. Gloeckler, G. & Fisk, L. A. Anisotropic beams of energetic particles upstream from the termination shock of the solar wind. *Astrophys. J.* **648,** L63–L66 (2006).
29. Giacalone, J. & Jokipii, J. R. Energetic particle intensities and anisotropies near the solar wind termination shock. *Astrophys. J.* **649,** L137–L140 (2006).
30. Jokipii, J. R., Giacalone, J. & Kóta, J. Transverse streaming anisotropies of charged particles accelerated at the solar wind termination shock. *Astrophys. J.* **611,** L141–L144 (2004).

Acknowledgements This work was supported by NASA (NAS7-03001).

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to E.C.S. (ecs@srl.caltech.edu).

LETTERS

Magnetic fields at the solar wind termination shock

L. F. Burlaga[1], N. F. Ness[2], M. H. Acuña[1], R. P. Lepping[1], J. E. P. Connerney[1] & J. D. Richardson[3]

A transition between the supersonic solar wind and the subsonic heliosheath was observed by Voyager 1, but the expected termination shock was not seen owing to a gap in the telemetry[1–4]. Here we report observations of the magnetic field structure and dynamics of the termination shock, made by Voyager 2 on 31 August–1 September 2007 at a distance of 83.7 AU from the Sun (1 AU is the Earth–Sun distance). A single crossing of the shock was expected, with a boundary that was stable on a timescale of several days. But the data reveal a complex, rippled, quasi-perpendicular supercritical magnetohydrodynamic shock of moderate strength undergoing reformation on a scale of a few hours. The observed structure suggests the importance of ionized interstellar atoms ('pickup protons') at the shock.

This Letter discusses three termination shock crossings (TS-2, TS-3 and TS-4) that were observed directly by Voyager 2 (Fig. 1). At least two additional termination shock crossings (TS-1 and TS-5) occurred when there were gaps in the telemetry. The multiple crossings imply motions of the termination shock, possibly caused by its large-scale motion or ripples propagating along it[5–7]. The times between crossings are of the order of magnitude expected for ripples propagating on the shock.

Observations by the magnetic field experiment on Voyager 2 (ref. 8) show no significant change in the direction of the magnetic field **B** across any of the three termination shock crossings in Fig. 1, indicating that the shock was quasi-perpendicular at each encounter. This geometry is necessary to provide the return of reflected ions from the shock[9,10], which ultimately produce some of the heating behind a high Mach number (supercritical) shock.

The structure of TS-3 is shown in Fig. 2. The foot, ramp and overshoot of TS-3 (ref. 9) moved past Voyager 2 in ~23 min, ~1.5 min and ~17 min, respectively. There is an increase in the magnetic field strength B in the foot, where **B** is compressed as the solar wind slows down[11]. The foot contains gyrating protons that are reflected from the shock[9,12,13]. A further abrupt decrease in speed to its lowest value occurs at the ramp, owing to the electric potential associated with cross-field currents and gradients in B in the ramp[14,15]. Plasma waves were observed at the ramp of TS-3 (ref. 16). The ratio of the maximum B in the overshoot to B downstream of the shock is ~2.5, indicating that the ratio of the magnetosonic Mach number M_{ms} to the critical fast Mach number ($M_c \approx 2$) is ~5 (ref. 17), suggesting that $M_{ms} \approx 10$ for TS-3, consistent with a supercritical shock.

From the passage time of the ramp (1.5 min) and the estimated shock speed (68 ± 17 km s^{-1}; ref. 11), the thickness of the ramp is ~6,000 km. The ramp size is usually expressed in terms of the ion inertial length $c/\omega_{pi,u}$, where c is the speed of light and $\omega_{pi,u}$ is the plasma frequency of an ion upstream. For TS-3, $c/\omega_{pi,u} \approx 6{,}000$ km and the thickness of the ramp is ~1 $c/\omega_{pi,u}$.

The strength of TS-3, measured by the ratio of the B in the solar wind upstream of the shock (B_1) to that behind the shock (B_2) is $B_2/B_1 = 1.7 \pm 0.1$. The corresponding density ratio is $N_2/N_1 = 1.4 \pm 0.2 \approx B_2/B_1$, consistent with a shock strength of 1.6 ± 0.2. The moderate strength of TS-3 might have been produced by the pickup protons and other suprathermal ions that were present[18], as considered in models[9,19–21].

The magnetic field strength oscillated quasi-periodically within the ramp (Fig. 3). The ramp moved past Voyager 2 in ~90 s, and each oscillation passed the spacecraft in ~12.3 s on average. The shock speed gives a characteristic length of the fluctuations in the ramp, $\lambda \approx 1{,}000$ km $= 0.15\ c/\omega_{pi,u}$. An oscillating ramp containing substructures with a scale of 0.2 $c/\omega_{pi,u}$ was observed at 1 AU (ref. 22), where shocks are not influenced by pickup protons.

A few hours after crossing TS-3, Voyager 2 crossed the termination shock (TS-4) back into the solar wind. The re-entry into the solar wind might have been caused by a ripple propagating on the shock's

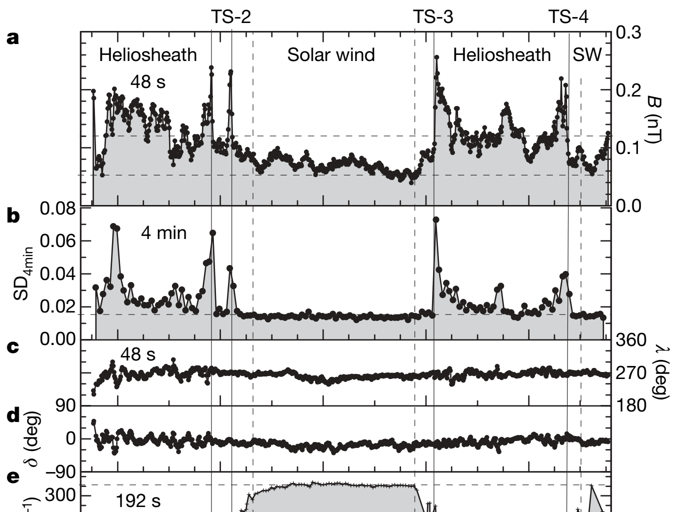


Figure 1 | Three crossings of the termination shock, illustrating reformation and the variability of its structure. The figure shows 48-s averages of magnetic field strength B (**a**) and the standard deviation of B for 4-minute intervals (**b**), azimuthal angle λ (**c**) and elevation angle δ (**d**), and the 192-s average of solar wind speed V (**e**) as a function of time measured in days from the beginning of 2007. (Here the directions λ and δ of **B** are in heliographic coordinates.) The time between the first and second of the observed termination shock crossings (TS-2 and TS-3, respectively) is ~3.7 h, and that between the second and third of the observed crossings (TS-3 and TS-4, respectively) is ~2.8 h. The Larmor frequency of a proton with the solar wind speed 300 km s^{-1} gyrating in the magnetic field upstream of the termination shock is $\Omega_{L,u}^{-1} \approx 17$ min. The two enhancements in B at TS-2 were separated in time by $\sim\Omega_{L,u}^{-1}$. The uncertainty in B is ± 0.03 nT.

[1]NASA/Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. [2]The Catholic University of America, Washington DC 20064, USA. [3]Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

surface[5,23]. The structure of TS-4 (Fig. 1) is different from that of TS-3. In the foot region, $B(t)$ (where t is time) appears as a narrow peak, and the overshoot is smaller than that of TS-3. Such a peak evolves from the foot as a result of bunching of reflected solar wind ions where they are turned back towards the shock[6,12]. The amplitude of the overshoot decreases while the step-like foot evolves to a peak. These changes were observed from TS-3 to TS-4 (Fig. 1). The structure of the termination shock evolved significantly within 2.7 h. The small peak in B at the front of TS-4 is expected to evolve to a new ramp on this timescale, as part of the shock reformation process[10,12,24,25].

Reformation of the local structure of a supercritical quasi-perpendicular shock was predicted by both hybrid and full particle simulations[10,13,25] for a shock with large M_{ms} and/or a low β, where β is the ratio of thermal pressure to magnetic pressure. Recall that we estimated that $M_{ms} \approx 10$ for TS-3. Neglecting pickup protons, $\beta = 0.04$ in the solar wind upstream of TS-3, so that β might be small even if the pickup protons contribute significantly to it. Reformation is a patchy cyclic shock reformation process with a characteristic time of the order of the downstream gyroperiod[25].

Further evidence for local reformation of the termination shock is provided by the qualitative transformation of the shock structure during the ~3.9-h interval between TS-2 and TS-3 (Fig. 1). At the front of TS-2, the bulk speed V increased continuously, rather than in a step-like form. Instead of a simple ramp-overshoot structure in B, there were two narrow enhancements in B resembling solitons, in each of which there was a change in B comparable to that in the ramp of TS-3. A shock structure at 1 AU resembling that of TS-2 was reported in fig. 5 of ref. 26.

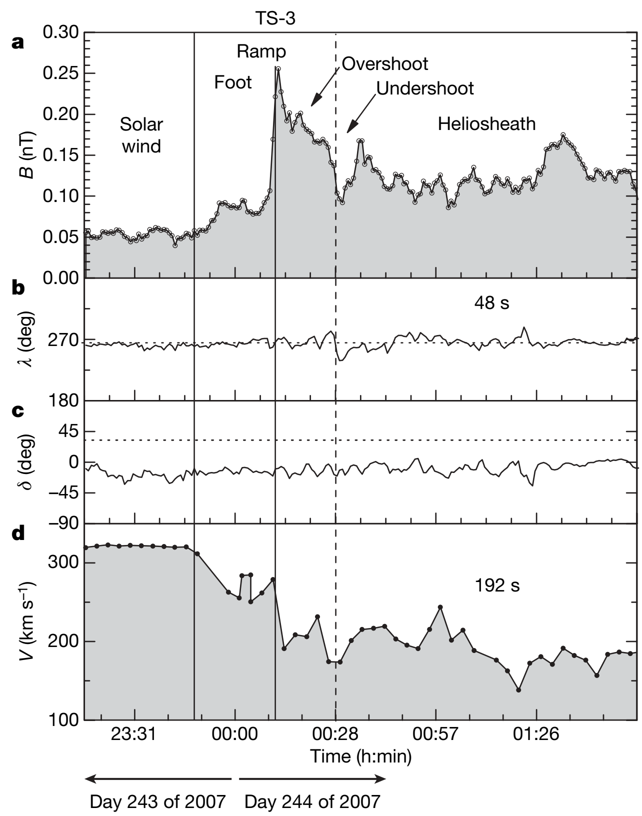


Figure 2 | TS-3 is a supercritical quasi-perpendicular shock. The 48-s averages of the magnetic field strength B (**a**) and its directions λ (**b**) and δ (**c**) are shown here together with the 192-s average of the solar wind speed V (**d**) across TS-3. The magnetic field strength profile shows the classical features of a supercritical quasi-perpendicular shock: a 'foot', 'ramp', 'overshoot', 'undershoot' and smaller oscillations, in that order.

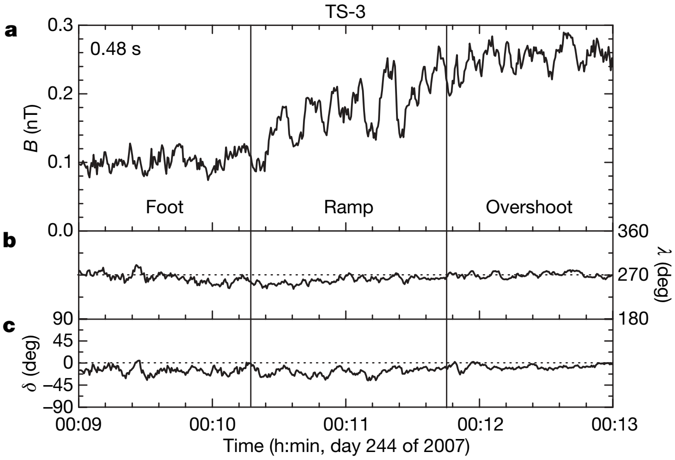


Figure 3 | The internal structure of the ramp of TS-3. The structure is based on observations of the magnetic field strength B (**a**) and its directions λ (**b**) and δ (**c**) at 0.48-s intervals. The magnetometer on Voyager 2 sampled the vector magnetic fields in the termination shock at a rate of 2.08 samples s^{-1}, and the spacecraft was able to transmit all of this information, making it possible to determine the complex internal structure of the ramp shown here.

Received 19 February; accepted 15 April 2008.

1. Burlaga, L. F. *et al.* Crossing the termination shock into the heliosheath: Magnetic fields. *Science* 309, 2027–2029 (2005).
2. Decker, R. B. *et al.* Voyager 1 in the foreshock, termination shock, and heliosheath. *Science* 309, 2020–2024 (2005).
3. Gurnett, D. A. & Kurth, W. S. Electron plasma oscillations upstream of the solar wind termination shock. *Science* 309, 2025–2027 (2005).
4. Stone, E. C. *et al.* Voyager 1 explores the termination shock region and the heliosheath beyond. *Science* 309, 2017–2020 (2005).
5. Winske, D. & Quest, K. B. Magnetic field and density fluctuations at perpendicular supercritical collisionless shocks. *J. Geophys. Res.* 93, 9681–9693 (1988).
6. Lembege, B. *et al.* Selected problems in collisionless shock physics. *Space Sci. Rev.* 110, 161–226 (2004).
7. Burgess, D. & Scholer, M. Shock front instability associated with reflected ions at the perpendicular shock. *Phys. Plasmas* 14, 012108 (2007).
8. Behannon, K. *et al.* Magnetic field experiment for Voyager-1 and Voyager-2. *Space Sci. Rev.* 21, 235–257 (1977).
9. Goodrich, C. C. in *Collisionless Shocks in the Heliosphere: Reviews of Current Research* (eds Tsurutani, B. T. & Stone, R. G.) 153–168 (Geophys. Monogr. Ser. Vol. 35, American Geophysical Union, Washington DC, 1985).
10. Scholer, M., Shinohara, M. I. & Matsukiyo, S. Quasi-perpendicular shocks: Length scale of the cross-shock potential, shock reformation, and implications for shock surfing. *J. Geophys. Res.* 108 (A1), doi:10.1029/2002JA009515 (2003).
11. Richardson, J. D., Kasper, J. C., Wang, C., Belcher, J. W. & Lazarus, A. J. Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock. *Nature* doi:10.1038/nature07024 (this issue).
12. Biskamp, D. & Welter, H. Numerical studies of magnetosonic collisionless shock waves. *Nucl. Fusion* 12, 663–666 (1972).
13. Phillips, P. E. & Robson, A. E. Influence of reflected ions on the magnetic structure of a collisionless shock. *Phys. Rev. Lett.* 29, 154–157 (1972).
14. Leroy, M. Structure of perpendicular shocks in collisionless plasmas. *Phys. Fluids* 26, 2742–2753 (1983).
15. Woods, L. C. On the structure of collisionless magneto plasma shock waves at supercritical Alfvén Mach numbers. *J. Plasma Phys.* 3, 435–442 (1969).
16. Gurnett, D. A. & Kurth, W. S. Intense plasma waves at and near the solar wind termination shock. *Nature* doi:10.1038/nature07023 (this issue).
17. Livesey, W. A. *et al.* ISEE 1 and 2 observations of magnetic field overshoots in quasi-perpendicular bow shocks. *Geophys. Res. Lett.* 9, 1037–1040 (1982).
18. Decker, R. B. *et al.* Mediation of the solar wind termination shock by non-thermal ions. *Nature* doi:10.1038/nature07030 (this issue).
19. Whang, Y. C., Burlaga, L. F. & Ness, N. F. Locations of the termination shock and heliopause. *J. Geophys. Res.* 100, 17015–17023 (1995).
20. Zank, G. P. *et al.* Interstellar pickup ions and quasi-perpendicular shocks: Implications for the termination shock and interplanetary shocks. *J. Geophys. Res.* 101, 457–477 (1996).
21. Lipatov, A. S. & Zank, G. P. Pickup ion acceleration at low β perpendicular shocks. *Phys. Rev. Lett.* 82, 3609–3612 (1999).
22. Newbury, J. A., Russell, C. T. & Gedalin, M. The ramp widths of high-Mach-number quasi-perpendicular collisionless shocks. *J. Geophys. Res.* 103 (A12), 29581–29593 (1998).

23. Lowe, R. E. & Burgess, D. The properties of rippling in quasi-perpendicular collisionless shock fronts. *Ann. Geophys.* **21,** 1–9 (2003).
24. Lembege, B. & Dawson, J. M. Self consistent study of a perpendicular collisionless and nonresistive shock. *Phys. Fluids* **30,** 1767–1788 (1987).
25. Lembege, B. & Savoini, P. Non-stationarity of a 2-D quasi-perpendicular supercritical collisionless shock by self-reformation. *Phys. Fluids* **4,** 3533–3548 (1992).
26. Bale, S. D. & Mozer, F. S. Measurement of large parallel and perpendicular electric fields on electron spatial scales in the terrestrial bow shock. *Phys. Rev. Lett.* **98,** 205001 (2007).

Acknowledgements We thank T. McClanahan and S. Kramer for support in the processing of the data. We also thank D. Berdischevsky for computing the instrument zero level corrections for the data in this paper, and for helping to solve the problems created by the erroneous decoding of a spacecraft systems command sent to Voyager 2 in 2006. N.F.N. was partially supported by a NASA grant to CUA.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.F.B (Leonard.F.Burlaga@nasa.gov).

LETTERS

Intense plasma waves at and near the solar wind termination shock

D. A. Gurnett[1] & W. S. Kurth[1]

Plasma waves are a characteristic feature of shocks in plasmas, and are produced by non-thermal particle distributions that develop in the shock transition layer. The electric fields of these waves have a key role in dissipating energy in the shock and driving the particle distributions back towards thermal equilibrium[1]. Here we report the detection of intense plasma-wave electric fields at the solar wind termination shock. The observations were obtained from the plasma-wave instrument on the Voyager 2 spacecraft[2]. The first evidence of the approach to the shock was the detection of upstream electron plasma oscillations on 1 August 2007 at a heliocentric radial distance of 83.4 AU (1 AU is the Earth–Sun distance). These narrowband oscillations continued intermittently for about a month until, starting on 31 August 2007 and ending on 1 September 2007, a series of intense bursts of broadband electrostatic waves signalled a series of crossings of the termination shock at a heliocentric radial distance of 83.7 AU. The spectrum of these waves is quantitatively similar to those observed at bow shocks upstream of Jupiter, Saturn, Uranus and Neptune.

The upstream plasma oscillations that provided the first indication that the spacecraft was approaching the termination shock were detected in the 311 Hz channel (Fig. 1) on 1 August 2007. These oscillations, known as Langmuir waves[3], are purely electrostatic oscillations that occur at a characteristic frequency known as the electron plasma frequency. The electron plasma frequency[4] is given by $f_p = 8{,}980\sqrt{n_e}$ Hz, where n_e is the local electron density in units of cm^{-3}. Although the electron plasma oscillations indicated that the spacecraft was close to the termination shock, because so little is known about the propagation of the electron beams responsible for these waves, it was difficult to estimate the distance to the shock or to determine how soon the shock crossing would occur.

For Voyager 1, comparable plasma oscillations were first detected about ten months before the spacecraft reached the termination shock[5]. Because Voyager 2 is moving away from the Sun slightly more slowly than Voyager 1 (3.2 AU yr^{-1} in comparison with 3.6 AU yr^{-1}), it might be expected to take slightly longer (perhaps $(3.6/3.2) \times 10 = 11.2$ months) to reach the shock. However, this simple estimate is complicated by the fact that the termination shock is not stationary. Its radial motion is controlled by the solar wind pressure, which varies considerably during the solar cycle. For the current phase of the solar cycle, which is near the solar minimum, the shock was expected to be moving inwards, owing to the low solar wind pressure. In contrast, for Voyager 1 the termination shock crossing occurred during the declining phase of the solar cycle, after a period of high solar wind pressure. Because of the higher solar wind pressure, it is believed that during this time the shock was moving outwards just beyond the spacecraft for a period of nearly two years[6,7]. It was during this relatively long period of close proximity to the shock that the upstream plasma oscillations were observed. Eventually, as the solar wind pressure began to decrease, the shock started to move inwards, leading to the shock crossing. On the basis of these considerations, we thought that the Voyager 2 crossing would occur much sooner than the Voyager 1 crossing, probably within several months and possibly within only a few weeks[8]. Nevertheless, it came as a surprise when, only 30 days after the first detection of the upstream

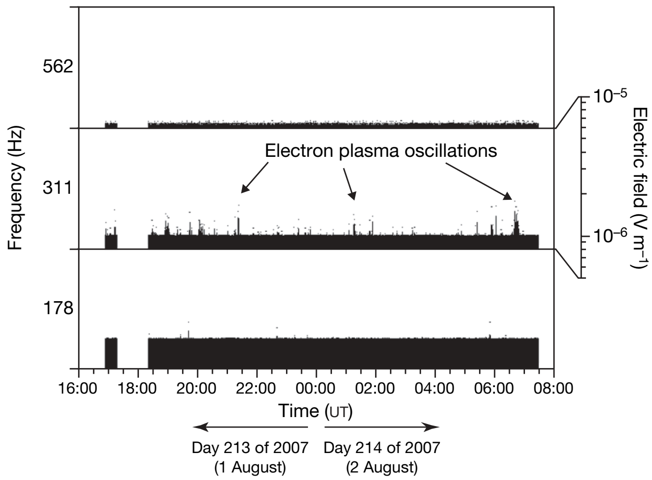


Figure 1 | The first detection of electron plasma oscillations by Voyager 2 upstream of the termination shock. Electric field intensities in the 178, 311 and 562 Hz channels of the plasma-wave instrument, at a heliocentric radial distance of 83.4 AU and heliographic longitudes and latitudes of respectively 288.6° and −31.6°. UT, universal time. The numerous impulsive intensity spikes in the 311 Hz channel are electron plasma oscillations. The frequency of these oscillations is consistent with the electron density ($n_e = 1.2 \times 10^{-3}\ \mathrm{cm}^{-3}$) expected in the solar wind at this heliocentric radial distance, and agrees with the plasma density measured by the plasma instrument at this time[10]. The emissions are very similar to those observed by Voyager 1 upstream of the termination shock[5], and indicate that the spacecraft is close to the termination shock. On the basis of observations ahead of planetary bow shocks, upstream waves of this type are known to be produced by beams of electrons that escape upstream along magnetic field lines that intersect the shock. According to the well-known theory of beam–plasma interactions, the oscillations are excited when the reduced electron velocity distribution function[1], $F(v)$, has a region of positive slope, $\mathrm{d}F/\mathrm{d}v > 0$. This condition is always satisfied ahead of the shock, owing to a time-of-flight mechanism that dictates that only those electrons with velocities greater than a well-defined cutoff velocity can reach a point in the upstream region[19]. Typical beam energies required to generate the upstream electron plasma oscillations are in the range from a few keV to several tens of keV. Although we have looked in the Voyager 2 plasma[10] and energetic charged particle[11] data for evidence of the electron beam responsible for the electron plasma oscillations, no clear correlation was found. The absence of an identifiable beam is probably due to the gap from about 6 to 35 keV between the electron energy coverage of the plasma instrument and that of the energetic-charged-particle instrument[20,21].

[1]Department of Physics and Astronomy, University of Iowa, Iowa City, Iowa 52242, USA.

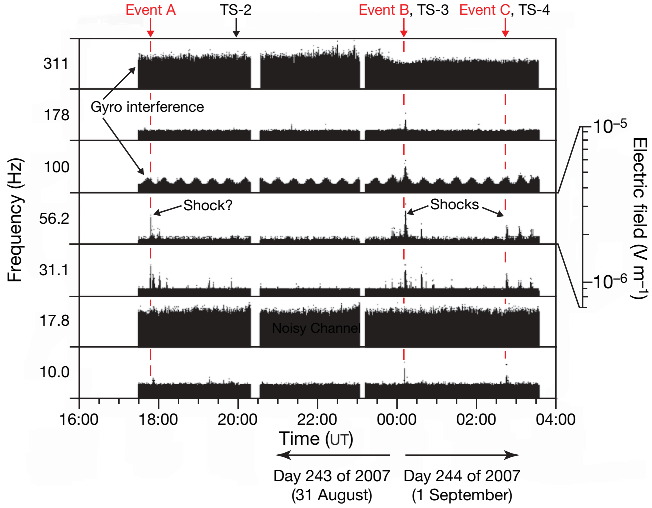


Figure 2 | Three bursts of broadband electrostatic noise of the type expected to occur at the termination shock. These bursts occurred during three time intervals: 17:50 to 18:15 UT on 31 August, and 00:11 to 00:13 UT and 02:45 to 03:35 UT on 1 September. The heliocentric radius is 83.67 AU, and the heliographic longitude and latitude are comparable to those in Fig. 1. The bursts (events A, B and C) and the shock crossings TS-2, TS-3 and TS-4, identified by the magnetometer and plasma teams[9,10], are indicated at the top of the plot. Events B and C correspond almost exactly to TS-3 and TS-4, respectively. Event A was not confirmed as a shock by the magnetometer and plasma teams, although there are variations in the magnetic field[9] around this time. No plasma waves were detected at TS-2.

plasma oscillations, a series of five closely spaced crossings of the termination shock were detected by the Voyager 2 particle and fields instruments[9–12] at a radial distance of only 83.7 AU, much closer than for the Voyager 1 termination shock crossing[6], which was at 94.1 AU. Because the Voyager 1 crossing occurred during a data gap[6], the Voyager 2 measurements provide the first direct observations of the structure of the termination shock.

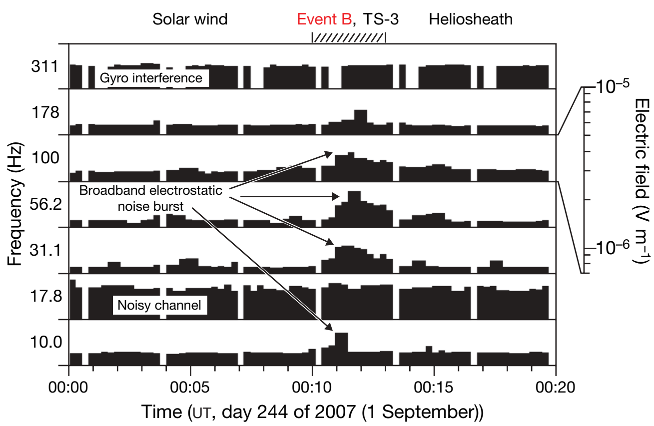


Figure 3 | A series of expanded timescale plots of the electric field intensities in the 10 to 311 Hz channels for event B. The heliocentric radius is 83.7 AU and the heliocentric longitude and latitude are comparable to those in Fig. 1. The broadband electrostatic noise burst that extends from 10 to 178 Hz between 00:11 and 00:13 UT corresponds almost exactly with the ramp in the magnetic field at TS-3 (ref. 9) and occurs just before an abrupt, well-defined peak in the plasma density[10]. The frequency of the electrostatic waves also is seen to increase systematically with increasing time as the spacecraft passes through the ramp, starting first in the 10 Hz channel at about 00:11 UT, progressing upwards through the 56.2 Hz channel about 30 seconds later, and finally reaching peak intensity in the 178 Hz channel at about 00:12 UT. This tendency, for the instability to start at long wavelengths (and low frequencies) and progress towards short wavelengths (and high frequencies), is characteristic of many electrostatic instabilities.

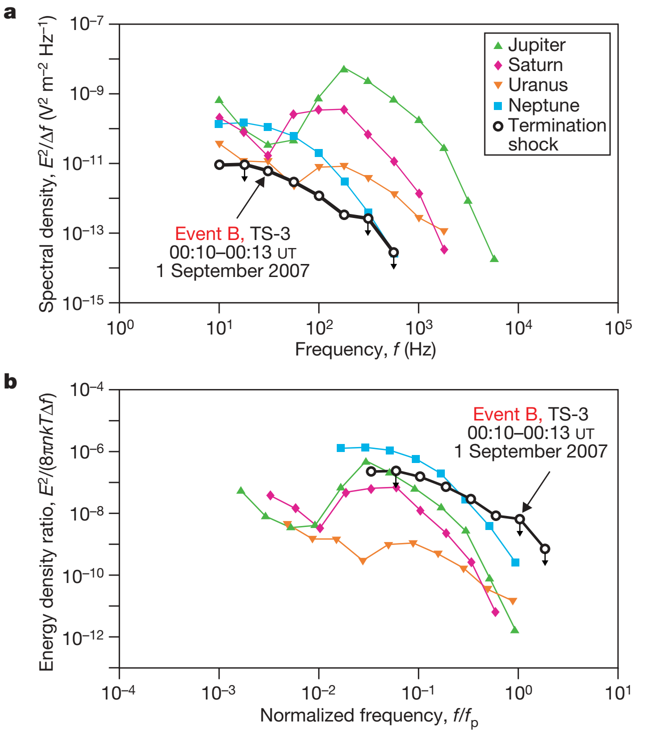


Figure 4 | Comparisons of the spectrum of the termination shock with spectra of planetary bow shocks. a, A comparison of the electric field spectrum at TS-3 with electric field spectra obtained by Voyager 1 and 2 at bow shocks upstream of Jupiter, Saturn, Uranus and Neptune[16]. The three downward arrows in the termination shock spectrum indicate upper limits, and Δf is the bandwidth used to convert the electric field amplitude, *E*, to spectral density. **b,** The same spectra normalized by dividing the electric field energy density, $E^2/8\pi$, by the plasma energy density, *nkT*, where *n* is the plasma density, *k* is Boltzmann's constant and *T* is the plasma temperature upstream of the shock, and by dividing the frequency, *f*, by the plasma frequency, f_p, upstream of the shock. The upstream parameters were taken to be $n = 1.2 \times 10^{-3}\ \mathrm{cm}^{-3}$, $T = 1.1 \times 10^4$ K and $f_p = 311$ Hz. This comparison shows that the normalized electric field intensities observed at the termination shock are in fact quite intense—more so than most of the planetary bow shock crossings, and comparably so to the bow shock crossing at Neptune.

The first indication in the Voyager 2 plasma-wave data of a possible crossing of the termination shock was the occurrence of three intense bursts of broadband electric field noise (labelled 'event A', 'event B' and 'event C' in Fig. 2) on 31 August and 1 September 2007. Events of this type are extremely rare and exhibit the intense broadband electrostatic wave spectrums that were expected at the termination shock[13]. Except for interplanetary shocks propagating outward from the Sun, which are generally less intense, no comparable broadband bursts have been observed in the Voyager 2 plasma-wave data since the crossing of Neptune's bow shock[14] on 24 August 1989. It is useful to compare these three events with measurements obtained from the particle and field instruments. Five successive in-and-out crossings of the termination shock, labelled TS-1 to TS-5, were identified by the magnetometer and plasma teams[9,10]. Of these, TS-1 and TS-5 occurred during data gaps, so no detailed comparisons can be made for these crossings. Of the remaining crossings, TS-3 and TS-4 occurred essentially simultaneously with events B and C (Fig. 2). Interestingly, no electric field bursts were observed during the shock crossing identified as TS-2. This shock crossing is unusual in that it consisted of two well-separated, soliton-like pulses[9] that are thought to be transient, non-dissipative structures that may be in the process of reforming into a shock farther upstream. For reasons that are not clearly understood, the characteristic signature of a shock was not present in the magnetometer and plasma data at the time of event A, the first broadband electrostatic burst.

TS-3 is especially clear and is worthy of further analysis. This shock, which represents a crossing from the solar wind into the heliosheath, is classified as a supercritical quasi-perpendicular shock[9,10]. High-resolution plots of the electric field intensities during this crossing (Fig. 3) show that the broadband burst of electric field noise (event B) starts at almost exactly the same time as the start of the steep ramp in the magnetic field, and ends at about the time that the magnetic field reaches peak intensity in the overshoot region[9]. A similar relationship also occurs at TS-4 (not shown). The close relationship between the electrostatic wave intensities and the magnetic field ramps observed at TS-3 and TS-4 implies that the electrostatic waves are closely related to the electrical currents that are responsible for the ramp in the magnetic field. Similar relationships between the broadband electrostatic wave intensity and the ramp in the magnetic field are commonly observed at planetary bow shocks[15,16].

Because the broadband electrostatic waves observed at the solar wind termination shock appear to be very similar to those observed at planetary bow shocks, it is interesting to compare the electric field spectrum observed at TS-3 with electric field spectra obtained by Voyagers 1 and 2 at bow shocks upstream of the magnetospheres of the outer planets[16]. This comparison (Fig. 4a) shows that the TS-3 spectrum has a shape that is very similar to those observed at planetary bow shocks, but with a distinct shift towards lower intensities and lower frequencies. This shift is a direct consequence of the $1/R^2$ decrease in the solar wind plasma density with increasing radial distance (R) from the Sun. If the electric field spectral densities are normalized by dividing by the solar wind energy density and the frequencies are normalized by dividing by the electron plasma frequency, then the TS-3 spectrum is found to be quite similar to the spectra observed at planetary bow shocks (Fig. 4b).

Although TS-3 has the classic electrostatic wave signature expected for a supercritical quasi-perpendicular shock, there are nevertheless large differences between successive crossings caused by fluctuations in the shock position. For example, TS-2 has no electrostatic wave signature at all, and event A, which has the classic electrostatic wave signature of a shock, was not identified as a shock by the magnetometer and plasma teams. Such complex variations are consistent with computer simulations of high-Mach-number supercritical shocks[17,18], which show that such shocks are highly unstable and continuously evolving.

Received 20 February; accepted 15 April 2008.

1. Gurnett, D. A. & Bhattacharjee, A. *Introduction to Plasma Physics* 277–278 (Cambridge Univ. Press, Cambridge, UK, 2005).
2. Scarf, F. L. & Gurnett, D. A. A plasma wave investigation for the Voyager mission. *Space Sci. Rev.* **21,** 289–308 (1977).
3. Tonks, L. & Langmuir, I. Oscillations in ionized gases. *Phys. Rev.* **33,** 195–210 (1929).
4. Stix, T. H. *The Theory of Plasma Waves* 10 (McGraw-Hill, New York, 1962).
5. Gurnett, D. A. & Kurth, W. S. Electron plasma oscillations upstream of the solar wind termination shock. *Science* **309,** 2025–2027 (2005).
6. Stone, E. C. *et al.* Voyager 1 explores the termination shock region and the heliosheath. *Science* **309,** 2017–2020 (2005).
7. Washimi, H. *et al.* A forecast of the heliospheric termination-shock position by three-dimensional MHD simuliations. *Astrophys. J.* **670,** L139–L142 (2007).
8. Kurth, W. S., Gurnett, D. A. & Decker, R. B. in *The Physics of the Inner Heliosheath: Voyager Observations, Theory, and Future Prospect (5th Annu. IGPP Internat. Astrophys. Conf.)* (eds Heerikhuisen, J., Florinski, V., Zank, G. P. & Pogorelov, N. V.) 116–121 (American Institute of Physics, Melville, New York, 2006).
9. Burlaga, L. F. *et al.* Magnetic fields at the solar wind termination shock. *Nature* doi:10.1038/nature07029 (this issue).
10. Richardson, J. D. *et al.* Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock. *Nature* doi:10.1038/nature07024 (this issue).
11. Decker, R. B. *et al.* Mediation of the solar wind termination shock by non-thermal ions. *Nature* doi:10.1038/nature07030 (this issue).
12. Stone, E. C. *et al.* An asymmetric solar wind termination shock. *Nature* doi:10.1038/nature07022 (this issue).
13. Kurth, W. S. & Gurnett, D. A. Plasma waves as indicators of the termination shock. *J. Geophys. Res.* **98,** 15129–15136 (1993).
14. Gurnett, D. A. *et al.* First plasma wave observations at Neptune. *Science* **246,** 1494–1498 (1989).
15. Fredricks, R. W., Kennel, C. F., Scarf, F. L., Crook, G. M. & Green, I. M. Detection of electric-field turbulence in the Earth's bow shock. *Phys. Rev. Lett.* **21,** 1761–1764 (1968).
16. Moses, S. L., Coroniti, F. V., Kennel, C. F., Kurth, W. S. & Gurnett, D. A. Comparison of plasma wave measurements in the bow shocks at Earth, Jupiter, Saturn, and Jupiter. *Geophys. Res. Lett.* **17,** 1653–1656 (1990).
17. Lembege, B. *et al.* Selected problems in collisionless-shock physics. *Space Sci. Rev.* **110,** 161–226 (2004).
18. Scholer, M., Shinohara, I. & Matsukiyo, S. Quasi-perpendicular shocks: Length scale of the cross-shock potential, shock reformation, and implication for shock surfing. *J. Geophys. Res.* **108,** doi:10.1029/2002JA009515 (2003).
19. Filbert, P. C. & Kellogg, P. J. Electrostatic noise at the plasma frequency beyond the Earth's bow shock. *J. Geophys. Res.* **84,** 1369–1381 (1979).
20. Bridge, H. S. *et al.* The plasma experiment on the 1977 Voyager mission. *Space Sci. Rev.* **21,** 259–287 (1977).
21. Krimigis, S. M. *et al.* The low energy charged particle (LECP) experiment on the Voyager spacecraft. *Space Sci. Rev.* **21,** 329–354 (1977).

Acknowledgements We wish to thank L. Burlaga and J. Richardson for discussions regarding the magnetometer and plasma data in advance of publication. The research at the University of Iowa was supported by NASA through the Jet Propulsion Laboratory.

Author Contributions D.A.G. is the principal investigator for the Voyager plasma-wave instruments and led the analysis. W.S.K. is the co-investigator and identified the wave signatures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to D.A.G. (donald-gurnett@uiowa.edu).

Vol 454 | 3 July 2008 | doi:10.1038/nature07068

LETTERS

Domination of heliosheath pressure by shock-accelerated pickup ions from observations of neutral atoms

Linghua Wang[1,2], Robert P. Lin[1,2], Davin E. Larson[2] & Janet G. Luhmann[2]

The solar wind blows an immense magnetic bubble, the heliosphere, in the local interstellar medium (mostly neutral gas) flowing by the Sun[1]. Recent measurements by Voyager 2 across the termination shock, where the solar wind is slowed to subsonic speeds before entering the heliosheath, found that the shocked solar wind plasma[2] contains only ~20 per cent of the energy released by the termination shock, whereas energetic particles[3] above ~28 keV contain only ~10 per cent; ~70 per cent of the energy is unaccounted for, leading to speculation[2,3] that the unmeasured pickup ions or energetic particles below 28 keV contain the missing energy. Here we report the detection and mapping of heliosheath energetic (~4–20 keV) neutral atoms produced by charge exchange of suprathermal ions with interstellar neutral atoms. The energetic neutral atoms come from a source ~60° wide in longitude straddling the direction of the local interstellar medium. Their energy spectra resemble those of solar wind pickup ions, but with a knee at ~11 keV instead of ~4 keV, indicating that their parent ions are pickup ions energized by the termination shock. These termination-shock-energized pickup ions contain the missing ~70 per cent of the energy dissipated in the termination shock, and they dominate the pressure in the heliosheath.

Energetic neutral atoms (ENAs) are remote tracers of energetic ion populations in distant regions[4], as they retain the parent ion's velocity in the charge exchange process. In the heliosheath, protons are the dominant ions, hydrogen is the dominant neutral species, and the proton–hydrogen atom cross-section is large, so heliosheath ENAs will be predominantly hydrogen. Furthermore, hydrogen ENAs above ~4 keV have a high probability (>95%) of surviving to the Earth's orbit[4].

ENAs were detected by the suprathermal electron (STE) sensors[5] of the IMPACT investigation[6] on the STEREO A and B spacecraft from June to October 2007 (Fig. 1). The ENA fluxes peaked near the flow direction of the local interstellar medium (ISM), with a width of ~40–60°. Although these ENAs are observed by different detectors on the two spacecraft at different times over more than ~4 months, the temporal profiles are closely similar (Fig. 2), indicating a distant, steady ENA source such as the heliosheath. The ENA fluxes observed by all the detectors show the same double-peak source structure, extending from ~230° to 290° ecliptic longitude (Fig. 3), with the maximum at ~245° and secondary peak at ~270°, straddling the interstellar gas source direction (254°). The observed longitudinal structure could be due, in part, to latitudinal variations (for example, tilted structures[7]).

The full 360° scan (~March 2007 to March 2008) detected no ENAs (upper limit of $\sim 2\times10^{-3}\,\mathrm{cm^{-2}\,s^{-1}\,sr^{-1}\,eV^{-1}}$ at 6.8 keV) from other directions, even though the termination shock should surround the entire heliosphere. For a tear-drop-shaped heliosphere and the wound-up upstream heliospheric magnetic field, a quasi-perpendicular termination shock (as observed by Voyager 2; ref. 2), which is more efficient than a quasi-parallel shock at accelerating suprathermal ions, would be found near the nose of the heliosphere relative to the interstellar gas flow. Previously reported detections of

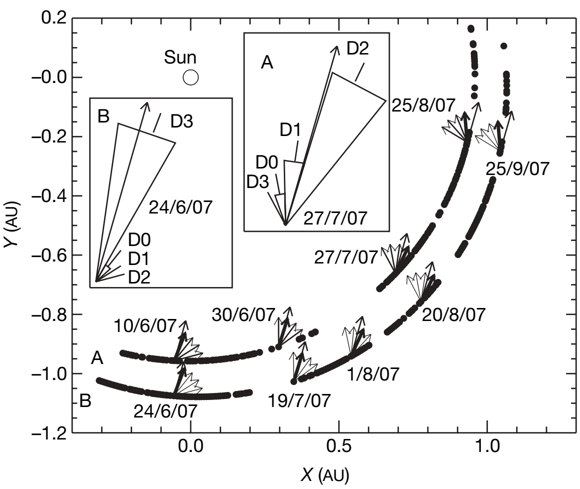


Figure 1 | The flow direction of energetic neutral atoms in the ecliptic plane. Main panel, filled circles indicate the location of STEREO A and B when ENAs were detected by the STE[5] sensors of the IMPACT investigation[6]. The gaps occur when the electron background is too high to distinguish ENAs. No ENAs are detected above background at other times in the year (March 2007 to March 2008) when the detectors are pointed in other directions, indicating that the ENA source is limited to ~60° ecliptic longitude around the nose of the heliosphere. For the specified dates, the particle flow direction is indicated by the short thick arrow for the detector observing the maximum ENA flux and by short thin arrows for the other three detectors, while the flow direction of the local ISM is shown by a long arrow. As both spacecraft orbit the Sun, the observed ENA fluxes peak, in turn, in the detector looking roughly towards the incoming local ISM flow. Insets, the anisotropy of 9 keV ENA fluxes observed on 24 June by STEREO B (left inset) and on 27 July by STEREO A (right inset). The latter shows that ENAs are detected by three adjacent detectors, indicating an angular width of ~40–60°. STE utilizes silicon semiconductor detectors with uniquely low thresholds to detect ≳2 keV electrons, and ≳4 keV protons/neutral hydrogen. STE-D has four detectors, D0, D1, D2, D3, observing particles coming towards the Sun. Each detector has a field of view covering from 40° S to 40° N in ecliptic latitude by 20° in ecliptic longitude; the fields of view of the four detectors together cover 80° longitude, as shown in the insets. ENAs are identified by their unchanging fluxes and flow direction when the interplanetary magnetic field changes direction (Supplementary Fig. 1), while beams of charged particles will vary with the interplanetary magnetic field.

[1]Physics Department, University of California, Berkeley, California 94720-7300, USA. [2]Space Sciences Laboratory, University of California, Berkeley, California 94720-7450, USA.

~0.2 to ~4–6 keV and >55 keV heliosheath ENAs indicate a peak from the tail[8–10] (opposite to the nose), with some evidence for ENAs from our secondary peak[11].

The measured hydrogen ENA energy spectra fit well to double power laws with a knee at ~10–12 keV (Figs 3 and 4), with low- and high-energy exponents of ~2–3 and ~4–7, respectively. The hydrogen ENA flux, j_{ENA}, is related to the average heliosheath energetic proton flux, j_{p}, (assumed isotropic) by:

$$j_{\mathrm{ENA}} = j_{\mathrm{p}}(\sigma_{\mathrm{pH}} n_{\mathrm{H}} + \sigma_{\mathrm{pHe}} n_{\mathrm{He}})L$$

with charge exchange cross-sections of proton–hydrogen and proton–helium respectively σ_{pH} and σ_{pHe}, source depth L, and densities[4,12,13] of neutral He and H respectively $n_{\mathrm{He}} = 0.015\ \mathrm{cm}^{-3}$ and $n_{\mathrm{H}} = 9n_{\mathrm{He}}$. The inferred spectrum of j_{p} for the major ENA peak (~60–70° in Fig. 3), together with measurements of ions with energies greater than ~40 keV (ref. 14) from Voyager 1's nearby crossing of the termination shock at ~94 AU (Fig. 4, upper trace) is a double power law with a knee at ~11 keV, plus a higher energy power-law tail. This is the same shape as the spectrum of the suprathermal ion population in the outer heliosphere[15], except there the knee is at ~4 keV (due to the maximum speed of twice the solar wind speed for pickup protons). This similarity indicates that the ENA parent heliosheath ions come from the energization of that population by the termination shock[15,16]. The ENA spectrum of the secondary peak is flatter than that of the major peak (Figs 3 and 4), consistent with the flatter spectrum of energetic, ≳28 keV ions observed by the nearby Voyager 2 crossing[3].

Voyager 2 (ref. 2) showed that the solar wind slows from ~310 to ~136 km s^{-1} in crossing the termination shock at ~84 AU, thus leading to a downstream energy of ~1.4 eV cm^{-3} (Supplementary Table 1). The solar wind plasma is heated only to ~10^5 K, so the downstream solar wind flow and thermal energy density (equivalent to the pressure) is only ~0.26 eV cm^{-3}. The observed power-law tail[3] of ≳28 keV ions contains ~0.1 eV cm^{-3}, so ~1 eV cm^{-3} is missing. The ~4–20 keV heliosheath protons (inferred from the ENAs) near the Voyager 2 termination shock crossing location (Fig. 3) have an energy density of ~0.1 eV cm^{-3}, but if these are termination-shock-heated pickup ions, their spectrum should extend down to solar wind energies (≲0.2 keV); if this is so, then their energy density is ~1 eV cm^{-3}, equal to the 'missing' energy, and their number density is ~1.8 × 10^{-3} cm^{-3}, ~75% of the solar wind.

Voyager 2 observed the solar wind speed to decrease in three steps, from ~400 km s^{-1} starting 0.7 AU upstream, to ~300 km s^{-1} at the termination shock[2], possibly due to the addition of ~30% to the solar wind density[17] in new pickup ions from ionization of interstellar neutrals by enhanced fluxes of energetic electrons[3] from the termination shock. Pickup ions accumulated previously should be a similar fraction[4], for a total close to 75%. Thus, most of the energy released by the termination shock is likely to go into heating the pickup protons[15], and these ions dominate the downstream pressure.

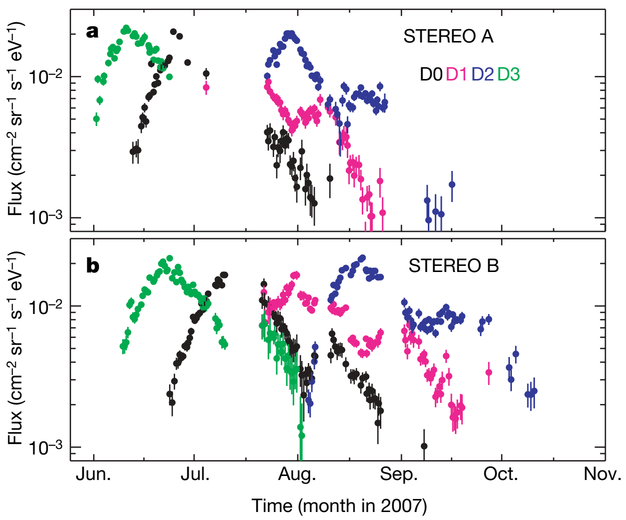


Figure 2 | The temporal variation of the 6.8 keV hydrogen ENA flux. Each filled circle indicates an average flux over a 2-h low-background period observed by STE-D (black, D0; pink, D1; blue, D2; green, D3). Each detector's field of view scans a full 360° ecliptic longitude during the year (March 2007 to March 2008). The temporal profile of the ENA flux is similar for all detectors; however, the flux peaks on 23 June in D3, 31 July in D1 and 20 August in D2 for STEREO B (**b**), while it peaks about 10–20 days earlier for STEREO A (**a**). No ENAs are detected above background at other times. In addition, the ENA flow is detected by three adjacent detectors during July–August and July–September for STEREO A and STEREO B, respectively, indicating a source region of ≳40–60° width in ecliptic longitude. Error bars, s.e.m. The fluxes below ~10^{-3} cm^{-2} sr^{-1} s^{-1} eV^{-1} are not statistically significant.

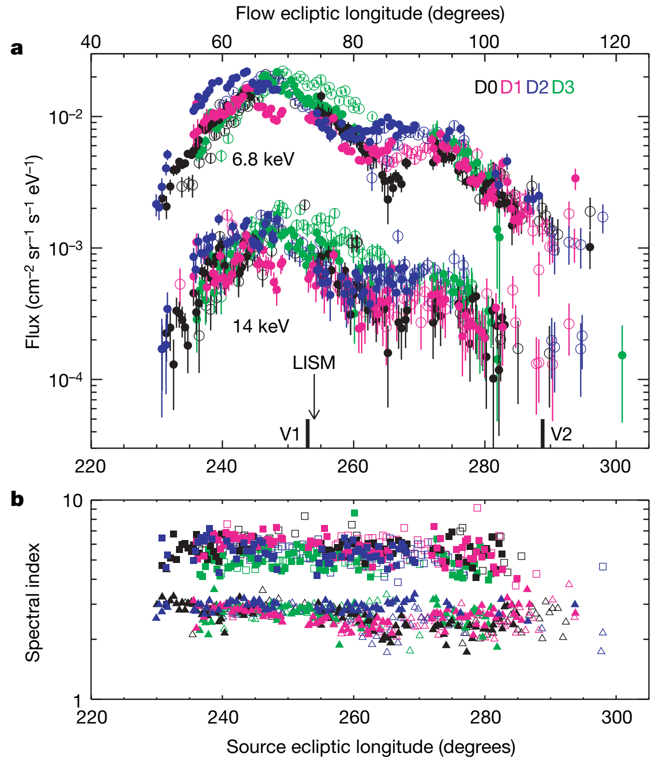


Figure 3 | The map of the heliosheath source of ENAs versus ecliptic longitude. Colour (black, pink, blue and green) indicates detectors (D0, D1, D2 and D3, respectively); open (filled) symbols show the data for STEREO A (B). **a**, The intensity of the ENA fluxes (cm^{-2} sr^{-1} s^{-1} eV^{-1}) at ~6.8 and 14 keV, plotted versus their source direction in ecliptic longitude (flow direction = source direction minus 180°, indicated at top). The arrow indicates the relative flow direction between the local ISM and Sun. The two short, thick vertical lines indicate the longitudes of the Voyager 1 (V1) and Voyager 2 (V2) crossings of the termination shock. Error bars, s.e.m.; LISM, local ISM. The ENA fluxes observed by different detectors show the same double-peak source structure as a function of ecliptic longitude from ~230° to 290°. No statistically significant ENA enhancement is detected at other ecliptic longitudes. The maximum occurs at ~240–250°, while the secondary, weaker peak is at ~270°. Each peak is ~20° full-width at half-maximum, ~30–40 AU wide, assuming the source is ~100 AU away. The differences between detectors may reflect temporal variations. **b**, The exponent below the 10–12 keV knee (triangles) and above (squares) for the double power-law fit to the ENA spectrum. For the stronger peak, the average power-law exponent is ~2.8 below the knee (10.9 keV) and ~5.6 above; for the weaker peak it is slightly flatter, ~2.5 below the break (10.7 keV) and ~5.4 above. This spectral difference may reflect the variability in space rather than in time, as it is seen by different detectors at different times.

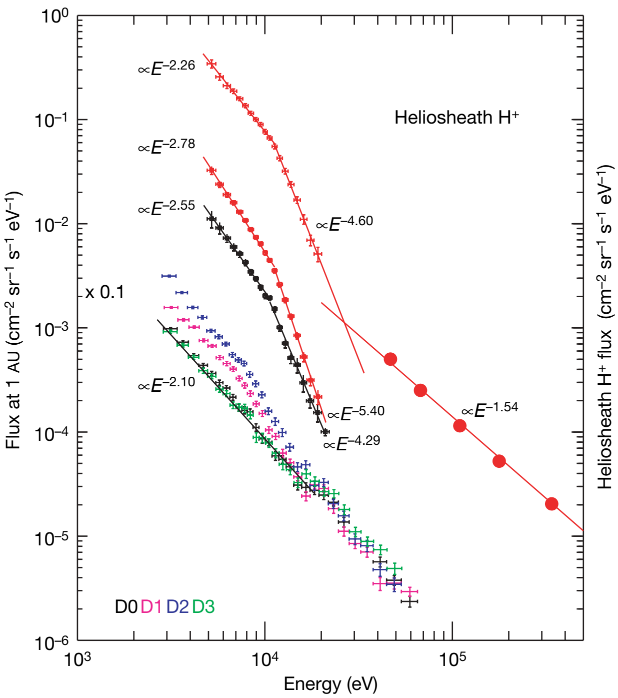


Figure 4 | Energy spectra of the particles. Bottom set of curves, STEREO B 9 September 2007 measured fluxes (divided by 10) versus energy (assuming electrons) in detectors D0, D1, D2, D3. D0 and D3 (black and green) measure the isotropic suprathermal electrons in the solar wind, well fitted (black line) by a power law with exponent ~2.1. The flattening around 20 keV is due to leakage from the onboard X-ray calibration source. The enhanced fluxes in the other two directions, D1 and D2 (pink and blue), at energies below ~15 keV are due to ENAs. Error bars here and in all curves represent s.e.m. Middle pair of curves, squares show ENA hydrogen spectra (red, maximum; black, secondary peak) at ~1 AU. These spectra are obtained by subtracting the electron background and correcting for the energy loss through the detector window (Supplementary Information). They fit to a double power law with exponents of ~2.8 and 2.6 below the knee (~11 keV) and of ~5.4 and 4.3 above, for the maximum (from near Voyager 1 crossing) and secondary peak (from near Voyager 2 crossing), respectively. Top curves, red symbols with error bars show the parent suprathermal proton fluxes (assumed isotropic) in the heliosheath, inferred from the ENA measurements at ~1 AU for the maximum. They fit to a double power law with exponent of 2.3 below the 11 keV knee and 4.6 above. We assume the distance to the ENA source is ~100 AU, and the source is 40 AU (equal to its FWHM width) deep. Red filled circles show the Voyager 1 measurements of the *in situ* heliosheath proton fluxes at energies above ~40 keV (ref. 14), after its termination shock crossing close to the ENA maximum location. Thus, the combination of these two spectra is representative of the heliosheath ion spectrum.

The ENA fluxes sometimes show significant differences between detectors (Fig. 3: longitude ranges of 235–240°, 245–248°, 253–270°), suggesting variability on a tens-of-days scale, perhaps from solar wind changes (but probably unrelated to the approximately hour-scale termination shock magnetic field variability[18]).

Received 16 March; accepted 1 May 2008.

1. Axford, W. I. *et al.* in *Solar Wind* (eds Sonett, C. P., Coleman, P. J. Jr & Wilcox, J. M.) 609–660 (Spec. Publ. 308, NASA, Washington DC, 1972).
2. Richardson, J. D., Kasper, J. C., Wang, C., Belcher, J. W. & Lazarus, A. J. Cool heliosheath plasma and deceleration of the upstream solar wind at the termination shock. *Nature* doi:10.1038/nature07024 (this issue).
3. Decker, R. B. *et al.* Mediation of the solar wind termination shock by non-thermal ions. *Nature* doi:10.1038/nature07030 (this issue).
4. Gruntman, M. *et al.* Energetic neutral atom imaging of the heliospheric boundary region. *J. Geophys. Res.* **106,** 15767–15781 (2001).
5. Lin, R. P. *et al.* The STEREO IMPACT Suprathermal Electron (STE) instrument. *Space Sci. Rev.* doi:10.1007/s11214-008-9330-7 (in the press).
6. Luhmann, J. G. *et al.* STEREO IMPACT investigation goals, measurements, and data products overview. *Space Sci. Rev.* doi:10.1007/s11214-007-9170-x (in the press).
7. Heerikhuisen, J. *et al.* The effects of global heliospheric asymmetries on energetic neutral atom sky maps. *Astrophys. J.* **655,** L53–L56 (2007).
8. Hilchenbach, M. *et al.* Detection of 55–80 keV hydrogen atoms of heliospheric origin by CELIAS/HSTOF on SOHO. *Astrophys. J.* **503,** 916–922 (1998).
9. Galli, A. *et al.* Direct measurements of energetic neutral hydrogen in the interplanetary medium. *Astrophys. J.* **644,** 1317–1325 (2006).
10. Wurz, P. *et al.* in *Physics of the Inner Heliosheath* (eds Heerikhuisen, J., Florinski, V., Zank, G. P. & Pogorelov, N. V.) 269–275 (AIP Conf. Proc. 858, AIP, New York, 2006).
11. Collier, M. R. *et al.* An unexplained 10–40° shift in the location of some diverse neutral atom data at 1 AU. *Adv. Space Res.* **34,** 166–171 (2004).
12. Witte, M. *et al.* Kinetic parameters of interstellar neutral helium: Final results from the ULYSSES/GAS Instrument. *Adv. Space Res.* **34,** 61–65 (2004).
13. Gloeckler, G. *et al.* Observations of the helium focusing cone with pickup ions. *Astron. Astrophys.* **426,** 845–854 (2004).
14. Decker, R. B. *et al.* in *Physics of the Inner Heliosheath* (eds Heerikhuisen, J., Florinski, V., Zank, G. P. & Pogorelov, N. V.) 73–78 (AIP Conf. Proc. 858, AIP, New York, 2006).
15. Gloeckler, G. & Fisk, L. A. in *Physics of the Inner Heliosheath* (eds Heerikhuisen, J., Florinski, V., Zank, G. P. & Pogorelov, N. V.) 153–158 (AIP Conf. Proc. 858, AIP, New York, 2006).
16. Fisk, L. A. *et al.* Acceleration of low-energy ions at the termination shock of the solar wind. *Astrophys. J.* **644,** 631–637 (2006).
17. Richardson, J. D., Paularena, K. I., Lazarus, A. J. & Belcher, J. W. Evidence for a solar wind slowdown in the outer heliosphere? *Geophys. Res. Lett.* **22,** 1469–1472 (1995).
18. Burlaga, L. F. *et al.* Magnetic fields at the solar wind termination shock. *Nature* doi:10.1038/nature07029 (this issue).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Mewaldt for suggestions, G. Gloeckler and E. Roelof for discussions, D. Curtis for instrument development and explanations of calibration data, and the Voyager team for sharing their results before publication. This research was supported in part by the National Aeronautics and Space Administration.

Author Contributions L.W. analysed the data and wrote the paper with direction and help from R.P.L. R.P.L. designed the Suprathermal Electron instrument and is the co-investigator in charge. D.E.L. assisted with the experimental work. J.G.L. is the principal investigator of the IMPACT investigation on STEREO that includes STE.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.W. (windsound@ssl.berkeley.edu).

LETTERS

Spatial cooperativity in soft glassy flows

J. Goyon[1], A. Colin[1], G. Ovarlez[2], A. Ajdari[3] & L. Bocquet[4,5]

Amorphous glassy materials of diverse nature—concentrated emulsions, granular materials, pastes, molecular glasses—display complex flow properties, intermediate between solid and liquid, which are at the root of their use in many applications[1–3]. A general feature of such systems, well documented yet not really understood, is the strongly nonlinear nature of the flow rule relating stresses and strain rates[4,5]. Here we use a microfluidic velocimetry technique to characterize the flow of thin layers of concentrated emulsions, confined in gaps of different thicknesses by surfaces of different roughnesses. We find evidence for finite-size effects in the flow behaviour and the absence of an intrinsic local flow rule. In contrast to the classical nonlinearities of the rheological behaviour of amorphous materials, we show that a rather simple non-local flow rule can account for all the velocity profiles. This non-locality of the dynamics is quantified by a length, characteristic of cooperativity within the flow at these scales, that is unobservable in the liquid state (lower emulsion concentrations) and that increases with concentration in the jammed state. Beyond its practical importance for applications involving thin layers (for example, coatings), these non-locality and cooperativity effects have parallels in the behaviour of other glassy, jammed and granular systems, suggesting a possible fundamental universality.

Glasses and jammed matter are at the heart of many challenging questions in condensed matter physics. The glass transition and the physical and mechanical properties of materials in the glassy or jammed state—from granular materials to dense colloidal suspensions—remain a puzzle to a large extent[2–8]. The transition itself is characterized by a dramatic slowing down of the system, associated with a dynamical arrest of the microscopic structure[3]. Although a large amount of work has been dedicated in recent years to demonstrating and characterizing dynamical heterogeneities in granular, glassy and glass-forming materials[9–16], the question of whether and how these dynamical heterogeneities influence the flow behaviour remains[17–19]. Indeed, one expects that such cooperative effects, which are in essence dynamical processes, should manifest themselves in the flow dynamics of jammed materials. This question is essential from the perspective of both fundamental science and applications: flows of jammed and glassy materials and glassy films are ubiquitous in nature and industry—for example, in granular flows, coatings with thin polymer films, lubrication processes in solid friction, in the food and cosmetic industries, and even in pedestrian dynamics[20].

In the present work, we aim to develop a constitutive law governing flows of jammed materials. We take advantage of a local velocity measurement technique to follow the local flow behaviour of a film of a confined soft glassy material (here a concentrated emulsion). We demonstrate the existence of finite-size effects in the flow dynamics that cannot be comprehended within our current understanding of the behaviour of fluids exhibiting yield properties.

To this end, we probe as sketched in Fig. 1 the flow of a jammed emulsion in various geometries and confinements: shear or pressure driven planar flow, in a wide gap Couette cell (centimetric gap) or a narrow microchannel (from tens to hundreds of micrometres in width). The emulsion is made of silicone droplets ($\sim$6.5 μm in diameter) in an index-matched glycerine–water mixture (Supplementary Methods). We are specifically interested in the local flow curves of the material, which relate the local shear stress σ to the local shear rate $\dot{\gamma}$ (Fig. 1c, d). Such curves can be constructed from the velocity profiles measured in the jammed emulsion in both geometries using local velocimetry techniques (Supplementary Methods; Fig. 1c, d insets). Indeed, in both geometries the stress distribution is known from mechanical equilibrium: $\sigma(R)=\Gamma/(2\pi HR^2)$ at a position R in the Couette cell (of height H) under a torque Γ; and $\sigma(z)=(\Delta P/L)[z-(w/2)]$ at a position $z-(w/2)$ from the centreline of the microchannel, under an applied pressure drop ΔP between the two ends of the channel of length L (ref. 21). Local shear rates $\dot{\gamma}$ are deduced directly from the velocity profiles. As shown in Fig. 1c, d, the local flow curves exhibit very different behaviours depending on the confinement of the material. In the wide gap Couette cell, all local flow curves obtained for different torques show a perfect superposition (Fig. 1c). The flow curve is furthermore well described by a classical Herschel-Bulkley model, $\sigma=\sigma_0+A\dot{\gamma}^{1/2}$, with σ_0 a 'dynamical' yield stress and A a material constant. On the other hand, in the narrow microchannels, data for different pressure drops ΔP are scattered in the whole figure and do not collapse on a single rheological curve (Fig. 1d; see also Supplementary Fig. 4). Accordingly, whereas a single Herschel–Bulkley model accounts perfectly for all the velocity profiles in the wide gap Couette cell (Fig. 1c inset), such is not the case for the narrow microchannels (Supplementary Fig. 8). Furthermore, below the jamming point, $\phi<\phi_c$ (with ϕ the volume fraction and ϕ_c its value at jamming), the local Herschel–Bulkley rheological model is fully able to reproduce the flow behaviour, even in narrow channels (Supplementary Fig. 3).

In summary, a key feature emerges from the experimental data: in the jammed state, $\phi>\phi_c$, there is no universal local relationship between stress and shear rate. The occurrence of finite-size effects in the flow properties of the jammed emulsion in narrow microchannels, and their absence in a wide geometry, point to the existence of extended spatial correlations, the effects of which increase as the volume of material decreases. As a further confirmation of this point, the overall shape of the velocity profiles, for the same wall shear stress, is shown to vary with the thickness of the channel (Fig. 2). Rougher surfaces are shown to induce higher shear rates for the same confinement and shear stress at the wall.

We did not find any evidence that the above results stem from a coupling to a structural change of the emulsion itself: we checked that neither the size distribution of the droplets, nor their shape, was affected by the flow. Moreover, in all geometries no variation of local droplet concentration, such as depletion or local ordering at the confining surfaces, was measured within the gap up to the experimental uncertainty of $\sim$1% in the volume fraction. This rules out

[1]LOF, Université Bordeaux 1, UMR CNRS-Rhodia-Bordeaux 1 5258, 33608 Pessac cedex, France. [2]Institut Navier, Université Paris Est, LMSGC, 77420 Champs sur Marne, France. [3]Gulliver, UMR CNRS – ESPCI 7083, 75005 Paris, France. [4]Laboratoire PMCN, Université Lyon 1, Université de Lyon, UMR CNRS 5586, 69622 Villeurbanne, France. [5]Physics Department, Technical University Munich, 85748 Garching, Germany.

density variations as the origin of non-local effects in the rheology (see Supplementary Fig. 10 and discussion in the legend). Such boundary effects on particle density have been reported in other disordered systems, such as granular materials and colloidal glasses[19,22], and their absence in the present case may be attributed to the relative softness of individual particles—which minimizes dilatancy effects—and the rather large polydispersity of the emulsions. We emphasize that the volume fraction of the emulsions, here typically 75–85%, is well above the jamming threshold (~64%), so that one does expect a relative insensitivity to local density variations. Altogether, our results suggest that the flow does not couple to a structural order parameter, so that non-local effects are 'intrinsic' to the rheology of the jammed emulsion, in line with similar observations in numerical simulations for glassy flows[17].

On the basis of these experimental facts, we now propose a theoretical framework able to rationalize these finite-size effects in the rheological behaviour. In concentrated emulsions, and more generally in soft glasses, flow occurs via a succession of reversible elastic deformations and local irreversible plastic rearrangements associated with a microscopic yield stress[23–26]. These localized plastic events induce a non-local, long ranged, elastic relaxation of the stress over the system. The number of plastic rearrangements per unit time, ν, plays the role of an inverse relaxation time and therefore controls the flow and relaxation of the material: a higher rate is associated with a more fluid system. Up to an elastic modulus G, the rate of rearrangements ν is equivalent to a 'fluidity', f, defined as $\sigma = (1/f)\dot{\gamma}$, a quantity that has been used in recent studies to characterize the rheology of fluids exhibiting yield properties[25]. In the absence of any non-local effect, the fluidity would reduce to its bulk value, $f_{\text{bulk}} = \dot{\gamma}/\sigma_{\text{bulk}}$ (note that for the Herschel-Bulkley expression reported above, $f_{\text{bulk}} \propto \dot{\gamma}$ in the quasistatic limit, as $\sigma \to \sigma_0$.) Now, owing to the non-local elastic relaxation of plastic rearrangements ('dynamical cooperativity')[24,26], a dynamically active region will induce agitation of its neighbours, and thus a higher rate of plastic rearrangements (and vice versa for a dynamically quiescent region). This suggests that non-local effects will occur for the rearrangement rates and consequently for the fluidity, f. In order to capture this physical picture, we assume that the local fluidity, $f(z)$, obeys a non-local equation in the steady state:

$$f(z) = f_{\text{bulk}} + \xi^2 \frac{\partial^2 f(z)}{\partial^2 z} \qquad (1)$$

In this equation, the non-local term describes how the plastic activity spreads spatially over the system due to non-local elastic relaxation: it accounts for the cooperativity involved in the plastic events occurring during flow, and ξ denotes a bulk 'flow cooperativity length'.

Together with the definition $f(z) = \dot{\gamma}(z)/\sigma(z)$, equation (1) constitutes our non-local rheological model for the emulsion. An analytical solution of equation (1) can be obtained, giving the fluidity at some position z in terms of an integral of the shear stress over the whole channel width. This solution requires two boundary conditions. A first is given by symmetry with respect to the centreline, and $\partial f/\partial z = 0$ at this point. A second boundary condition for the fluidity results from the dynamical behaviour at the surfaces. As we describe in the Supplementary Information, we did characterize the surface flow behaviour by investigating the relationship between wall shear rate and wall shear stress, whose ratio defines the surface fluidity $f_{\text{wall}} = \dot{\gamma}_{\text{wall}}/\sigma_{\text{wall}}$. Interestingly, these measurements demonstrate the existence of a specific surface rheology, $f_{\text{wall}} = f_{\text{wall}}(\dot{\gamma}_{\text{wall}})$, which for a given emulsion is found to depend only on the surface roughness (Supplementary Methods, Supplementary Fig. 7). We note that whatever the boundary condition at the wall, the solution of equation (1) can not reduce to its local form, $f_{\text{bulk}}(z) = f_{\text{bulk}}(\sigma(z))$, owing to the diffusive, non-local term. For convenience we then solved equation (1) numerically with the two above boundary conditions, and

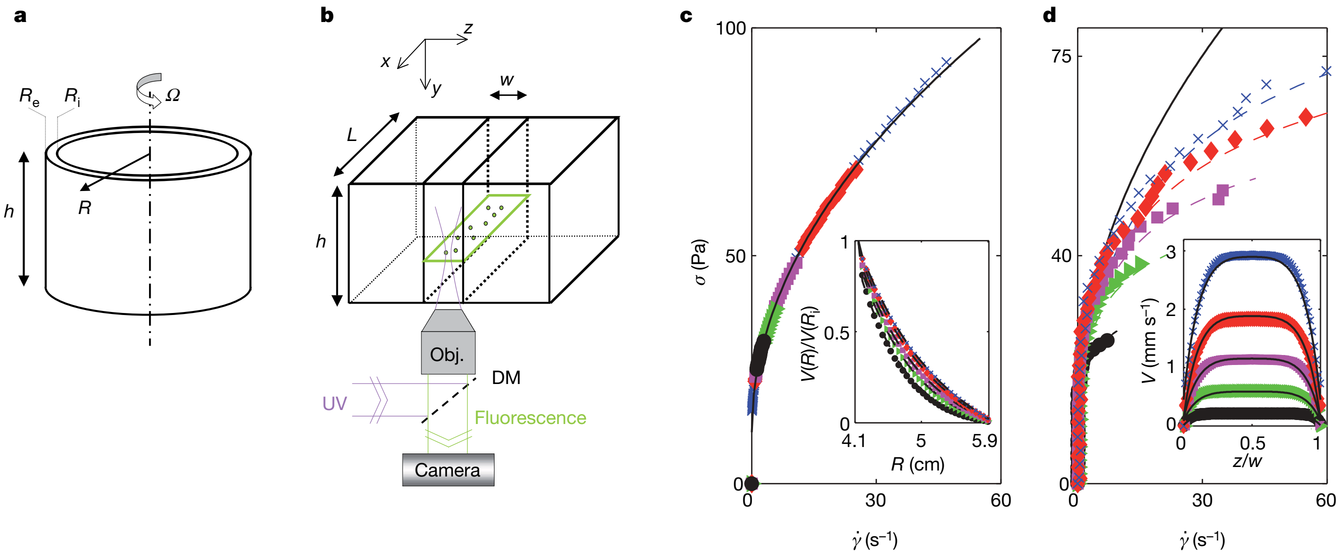


Figure 1 | Local flow curves measured by magnetic resonance imaging and microfluidic velocimetry. **a**, Schematic of the wide gap Couette cell (inner and outer cylinder radii $R_i = 4.1$ cm and $R_e = 5.9$ cm, respectively, gap $w = R_e - R_i = 1.8$ cm, height $h = 11$ cm). **b**, Flow in a microchannel and particle imaging velocimetry. DM, dichroic mirror; UV, ultraviolet lamp; Obj., microscope objective. Flow direction is along x. **c**, **d**, Local flow curves extracted from the measurements of the velocity profiles in both geometries for an emulsion with volume fraction $\phi = 0.75$ and 20% polydispersity. The shear rates are obtained by finite differencing of the velocity profiles given in the insets. Different colours correspond to different rotation velocities (**c**) or different pressure drops (**d**). In **c** are shown local flow curves, $\sigma(z)$ versus $\dot{\gamma}(z)$, extracted from the velocity profiles measured within a wide gap Couette geometry (see inset). The solid line corresponds to the Herschel–Bulkley model with $\sigma_0 = 11.6$ Pa, $A = 11.2$ Pa s$^{1/2}$. Inset, Dimensionless velocity profiles $V(R)/V(R_i)$ as a function of the radial coordinate R. From bottom to top the rotation velocities are 5, 10, 20, 50 and 100 r.p.m. In **d** are shown local flow curves, $\sigma(z)$ versus $\dot{\gamma}(z)$, extracted from the velocity profiles measured in a $w = 250$ μm thick microchannel with rough surfaces, for various pressure drops as a function of the reduced coordinate z/w (see inset). No overlap of the local flow curves is observed. Dashed lines are predictions for the local flow curves at the given ΔP, as obtained from the non-local rheological model of equation (1) with a flow cooperativity length $\xi = 22.3$ μm. The solid line corresponds to the Herschel–Bulkley model with $\sigma_0 = 11.6$ Pa, $A = 11.2$ Pa s$^{1/2}$. Inset, corresponding velocity profiles measured for ΔP equal to 300, 450, 600, 750, 900 mbar. The length of the channel is $L = 14$ cm. Solid lines are the velocity profiles predicted by the non-local rheological model of equation (1). Slippage, which is found to occur at the surfaces as detailed in Supplementary Fig. 7, does not affect the resulting local flow curves as plotted in the main figure.

velocity profiles were deduced by integrating the computed local shear rate $\dot{\gamma}(z)$.

The only remaining parameter is the bulk 'flow cooperativity length' ξ in equation (1). A crucial and surprising finding is that a unique (constant) length ξ accounts for all experimental data for the flow profiles and local flow curves for a given emulsion (Figs 1d and 2), independently of the pressure drop, confinement and surface nature (rough or smooth). That so much data can be fitted using a single and constant length ξ is quite remarkable. Typical comparisons are shown in Figs 1d and 2 for both the velocity profiles and the corresponding local flow curves (see also Supplementary Figs 5 and 6). We mention that alternative non-local rheological models, involving, for example, non-local terms in the flow curve[27], are unable to reproduce all measured flow profiles (Supplementary Fig. 9).

Furthermore, a similar, very good agreement has been obtained for all (jammed) emulsions investigated, with various volume fractions and polydispersities (Supplementary Figs 5 and 6). Figure 3 reports the variation of the measured flow cooperativity length ξ as a function of the volume fraction for emulsions with two different dispersities. Below the jamming concentration $\phi<\phi_c$, we could not measure finite-size effects in the flow profiles within the experimental uncertainty (Supplementary Fig. 3), so that $\xi\approx 0$ in this case. A strong increase of ξ is observed above the jamming threshold, in line with the appearance of elastic and yield properties of the materials in this regime.

Non-local effects in the jammed state are expected to affect the flow behaviour as the cooperativity length ξ is comparable to the confinement width w, that is $\xi/w=\mathcal{O}(1)$. According to equation (1), such confinement effects arise due to fluidity gradients, which may have their origin either in boundary effects or in the existence of gradients of stress occurring at a scale comparable to ξ, that is, $(\nabla\sigma)/\sigma\approx\xi^{-1}$. This is the case for the microfluidic configurations: the orders of magnitude of boundary effects and stress gradients are given by ξ/w, which is indeed up to $\sim$0.5 in this geometry (Figs 1d and 2). On the other hand, for the wide Couette cell (Fig. 1c), boundary effects and stress gradients are much weaker (as estimated by $\xi/w\approx 10^{-3}\ll 1$ and $\xi(\nabla\sigma)/\sigma\approx 2\xi/R_i\approx 10^{-3}\ll 1$). This justifies *a posteriori* the absence of any measured non-local effects in this cell.

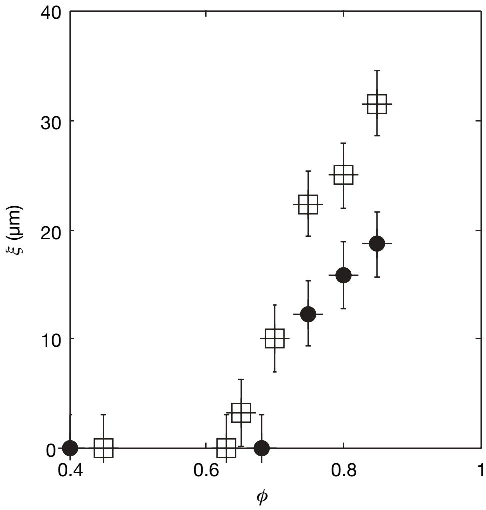


Figure 3 | Flow cooperativity length ξ as a function of emulsion volume fraction ϕ. Squares and circles correspond to emulsions with respective polydispersities of 20% and 36%. The corresponding jamming volume fractions, as determined by the occurrence of a non-vanishing yield stress, are $\phi_c = 0.64$ (squares) and $\phi_c = 0.68$ (circles). Error bars denote the dispersion of the experimental values over various experiments, $\pm$1 s.d.

Furthermore, as the confinement ratio ξ/w goes to zero, local flow curves increasingly follow the bulk Herschel-Bulkley prediction (Supplementary Fig. 11).

Non-local effects in the flow properties may be *a priori* expected at the microscopic scale[27]. However, we measure a 'flow cooperativity length' only in the jammed state, providing a direct and novel connection between non-local effects and the jamming transition. Furthermore, we emphasize that this length is associated with correlations of a dynamical quantity (the fluidity, that is, rearrangements) and not with static structural order. Typically for the volume fraction range studied here, ξ varies between 0 and 30 μm, compared with the droplet size $d\approx 6$ μm, giving a cooperatively rearranging region involving up to a few hundred particles. These values are close to

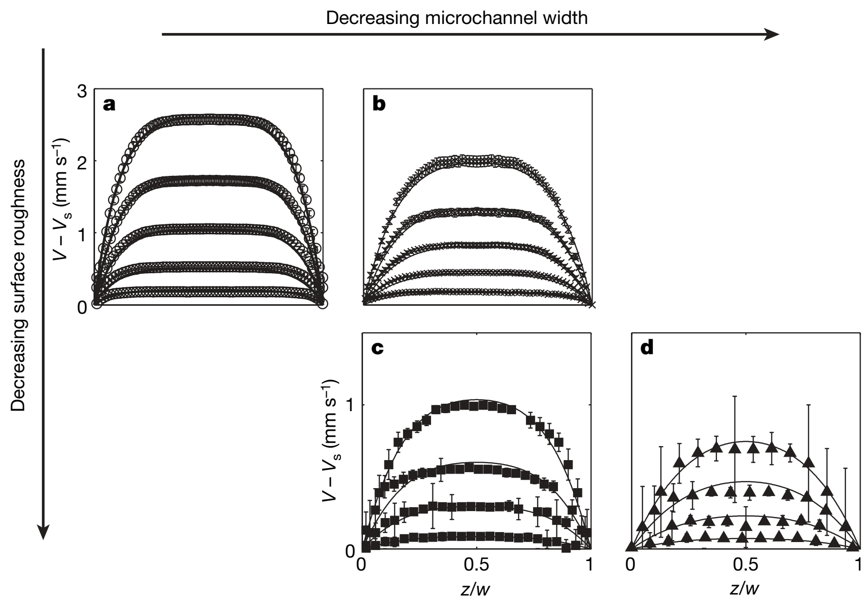


Figure 2 | Velocity profiles in microchannels with different widths w and confining wall roughness, corrected for the slip velocity. **a**, Rough microchannel with $w = 250$ μm; **b**, rough microchannel with $w = 125$ μm; **c**, smooth microchannel with $w = 112$ μm; and **d**, smooth microchannel with $w = 56$ μm. Plotting symbols show data points; solid lines show velocity profiles predicted by the non-local model of equation (1) with a flow cooperativity length $\xi = 22.3$ μm. The various experimental velocity profiles correspond to different pressure drops ΔP, tuned to get the same range of wall shear stress σ_{wall} in the various geometries. From bottom to top in the indicated panels, the values of σ_{wall} (Pa) are: **a**, 27, 41, 55, 68, 82; **b**, 30, 45, 60, 75, 90; **c**, 45, 60, 75, 91; **d**, 48, 65, 82, 97 Pa. The volume fraction is $\phi = 75\%$, with 20% polydispersity. Error bars, $\pm$1s.d. ($n = 450$).

estimates for the dynamical heterogeneities measured in glassy materials[10,11] or in jammed granular materials[14,16,28]. Furthermore, similar observations of non-locality have been reported in granular flows close to the jamming transition[15,29], suggesting further universal characteristics.

However, we emphasize that the present length characterizing flow cooperativity differs fundamentally from that which characterizes dynamical heterogeneities involved in spontaneous fluctuations. Indeed, whereas these dynamical heterogeneities have been measured to exhibit a maximum at the jamming point[10,30], the present flow cooperativity length ξ is non-vanishing in the jammed region only, and increases when going deeper in the jammed phase. This suggests that the flow behaviour of glassy and jammed materials involves physical mechanisms distinct from those associated with the glass transition. We expect that the flow cooperativity length ξ measures the zone of influence of localized plastic events occurring during flow and not the size of mobile regions. Our results thus provide a novel perspective for developing a theoretical framework to describe the non-local flow behaviour of glassy and jammed systems, such as soft glasses and granular materials.

Received 5 November 2007; accepted 18 April 2008.

1. Jop, P., Forterre, Y. & Pouliquen, O. A constitutive law for dense granular flows. *Nature* **441**, 727–730 (2006).
2. Debenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **410**, 259–267 (2001).
3. Biroli, G. A new kind of phase transition? *Nature Phys.* **3**, 222–223 (2007).
4. Coussot, P. Rheophysics of pastes: a review of microscopic modelling approaches. *Soft Matter* **3**, 528–540 (2007).
5. Fuchs, M. & Cates, M. E. Theory of nonlinear rheology and yielding of dense colloidal suspensions. *Phys. Rev. Lett.* **89**, 248304 (2002).
6. Goldenberg, C. & Goldhirsch, I. Friction enhances elasticity in granular solids. *Nature* **435**, 188–191 (2005).
7. Ostojic, S., Somfai, E. & Nienhuis, B. Scale invariance and universality of force networks in static granular matter. *Nature* **439**, 828–830 (2006).
8. Liu, A. J. & Nagel, S. R. Jamming is not just cool any more. *Nature* **396**, 21–22 (1998).
9. Berthier, L. Time and length scales in supercooled liquids. *Phys. Rev. E* **69**, 020201(R) (2004).
10. Weeks, E. R., Crocker, J. C., Levitt, A. C., Schofield, A. & Weitz, D. A. Three dimensional direct imaging of structural relaxation near the colloidal glass transition. *Science* **287**, 627–631 (2000).
11. Berthier, L. *et al.* Direct experimental evidence of a growing length scale accompanying the glass transition. *Science* **310**, 1797–1800 (2005).
12. Gao, Y. & Kilfoil, M. L. Direct imaging of dynamical heterogeneities near the colloid-gel transition. *Phys. Rev. Lett.* **99**, 078301 (2007).
13. Priestley, R. D., Ellison, C. L., Broadbelt, L. J. & Torkelson, J. M. Structural relaxation of polymer glasses at surfaces, interfaces and in between. *Science* **309**, 456–459 (2005).
14. Dauchot, O., Marty, G. & Biroli, G. Dynamical heterogeneity close to the jamming transition in a sheared granular material. *Phys. Rev. Lett.* **95**, 265701 (2005).
15. Pouliquen, O. Velocity correlations in dense granular flows. *Phys. Rev. Lett.* **93**, 248001 (2004).
16. Keys, A. S., Abate, A. R., Glotzer, S. C. & Durian, D. J. Measurement of growing dynamical length scales and prediction of the jamming transition in a granular material. *Nature Phys.* **3**, 260–264 (2007).
17. Varnik, F., Bocquet, L., Barrat, J.-L. & Berthier, L. Shear localization in a model glass. *Phys. Rev. Lett.* **90**, 095702 (2003).
18. Leonforte, F., Tanguy, A., Wittmer, J. P. & Barrat, J.-L. Inhomogeneous elastic response of silica glass. *Phys. Rev. Lett.* **97**, 055501 (2006).
19. Isa, L., Besseling, R. & Poon, W. C. Shear zones and wall slip in the capillary flow of concentrated colloidal suspensions. *Phys. Rev. Lett.* **98**, 198305 (2007).
20. Helbing, D., Farkas, I. & Vicsek, T. Simulating dynamical features of escape panic. *Nature* **407**, 487–490 (2000).
21. Degré, G. *et al.* Rheology of complex fluids by particle image velocimetry in microchannels. *Appl. Phys. Lett.* **89**, 024104 (2006).
22. Mueth, D. M. *et al.* Signatures of granular microstructure in dense shear flows. *Nature* **406**, 385–389 (2000).
23. Princen, H. M. Rheology of foams and highly concentrated emulsions. *J. Colloid Interface Sci.* **91**, 160–175 (1983).
24. Schall, P., Weitz, D. A. & Spaepen, F. Structural rearrangements that govern flow in colloidal glasses. *Science* **318**, 1895 (2007).
25. Picard, G., Ajdari, A., Bocquet, L. & Lequeux, F. A simple model for heterogeneous flows of yield stress fluids. *Phys. Rev. E* **66**, 051501 (2002).
26. Picard, G., Ajdari, A., Lequeux, F. & Bocquet, L. Slow flows of yield stress fluids: complex spatiotemporal behavior within a simple elastoplastic model. *Phys. Rev. E* **71**, 010501(R) (2005).
27. Dhont, J. K. G. A constitutive relation describing the shear-banding transition. *Phys. Rev. E* **60**, 4534–4544 (1999).
28. Kamrin, K. & Bazant, M. Z. A stochastic flow rule for granular materials. *Phys. Rev. E* **75**, 041301 (2007).
29. Deboeuf, S., Lajeunesse, E., Dauchot, O. & Andreotti, B. Flow rule, self-channelization and levees in unconfined granular flows. *Phys. Rev. Lett.* **97**, 158303 (2006).
30. Goldman, D. I. & Swinney, H. L. Signatures of glass formation in a fluidized bed of hard spheres. *Phys. Rev. Lett.* **96**, 145702 (2006).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Discussions with J.-L. Barrat and B. Andreotti are acknowledged. This project was supported by Rhodia, Région Aquitaine and the ANR. L.B. acknowledges support from the von Humboldt Foundation.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to L.B. (lyderic.bocquet@univ-lyon1.fr) and A.C. (annie.colin-exterieur@eu.rhodia.com).

LETTERS

Vibrational excitation through tug-of-war inelastic collisions

Stuart J. Greaves[1]*, Eckart Wrede[2], Noah T. Goldberg[3]*, Jianyang Zhang[3], Daniel J. Miller[3] & Richard N. Zare[3]

Vibrationally inelastic scattering is a fundamental collision process that converts some of the kinetic energy of the colliding partners into vibrational excitation[1,2]. The conventional wisdom is that collisions with high impact parameters (where the partners only 'graze' each other) are forward scattered and essentially elastic, whereas collisions with low impact parameters transfer a large amount of energy into vibrations and are mainly back scattered[3]. Here we report experimental observations of exactly the opposite behaviour for the simplest and most studied of all neutral–neutral collisions: we find that the inelastic scattering process $H + D_2(v = 0, j = 0, 2) \rightarrow H + D_2(v' = 3, j' = 0, 2, 4, 6, 8)$ leads dominantly to forward scattering (v and j respectively refer to the vibrational and rotational quantum numbers of the D_2 molecule). Quasi-classical trajectory calculations show that the vibrational excitation is caused by extension, not compression, of the D–D bond through interaction with the passing H atom. However, the H–D interaction never becomes strong enough for capture of the H atom before it departs with diminished kinetic energy; that is, the inelastic scattering process is essentially a frustrated reaction in which the collision typically excites the outward-going half of the H–D–D symmetric stretch before the $H–D_2$ complex dissociates. We suggest that this 'tug of war' between H and D_2 is a new mechanism for vibrational excitation that should play a role in all neutral–neutral collisions where strong attraction can develop between the collision partners.

Vibrationally inelastic scattering is usually treated by considering the repulsive part of the potential and it is assumed to happen by means of impulsive compression of one of the bonds of the collision partners. This behaviour well describes the vibrational excitation of targets with closed-shell configurations, such as $He + H_2$ (ref. 4). It has also been shown to provide an explanation for what is observed in collisions of fast H atoms with CO (ref. 5) and CO_2 (ref. 6). Sometimes attractive forces between the collision partners have been invoked to explain increased vibrational transfer. An example is the study of H + NO (ref. 7) in which it was suggested that vibrational excitation is enhanced by the acceleration of the incoming H atom as it passes through the well before it strikes the NO molecule and compresses the NO bond. In general, however, the attractive portion of the potential seems to play an auxiliary role in accounting for vibrational excitation when the collision energy is large in comparison with the well depth[8]. We examine the collision of fast H atoms (1.72 eV) with supersonically cooled D_2 molecules in which the collision energy greatly exceeds the van der Waals well depth (~3 meV) for the collinear geometry. We find that H–D attraction nevertheless plays an important role in the production of $D_2(v' = 3)$, as will be explained in what follows.

The experimental setup is virtually identical to that described in refs 9–11. A mixture of 3% HBr in D_2 with a typical backing pressure of 1.3 bar is introduced through a 10-Hz pulsed valve into a vacuum chamber. The reactants undergo internal and translational cooling in the supersonic expansion, and the D_2 is prepared almost equally in the $v = 0, j = 0$ and $v = 0, j = 2$ states. Because of symmetry (causing the preservation of ortho and para states during collision), $D_2(v = 0, j = 1, 3)$ molecules do not contribute to what is observed. Two linearly polarized, tuneable ultraviolet laser pulses ($\Delta\tau \approx 5$ ns) intersect the molecular beam at right angles. The reaction is initiated by photolysing HBr with a pulse of laser light ($\lambda \approx 209.4$ nm, $E_{pulse} \approx 300\ \mu J$) to produce monoenergetic H atoms with a well-defined spatial anisotropy. Single collisions with H atoms cause some of the initially cold D_2 molecules to become rovibrationally excited; nascent $D_2(v' = 3, j')$ is state-selectively probed via (2+1) resonance-enhanced multiphoton ionization on the Q branch lines of the (0, 3) $E,F^1\Sigma_g{}^+–X^1\Sigma_g{}^+$ electronic band system, where E,F refers to the double-well upper electronic state ($220.1 \le \lambda \le 221.9$ nm, $E_{pulse} \approx 300\ \mu J$). The resultant $D_2{}^+$ ions are formed in the extraction region of a Wiley–McLaren time-of-flight mass spectrometer and are accelerated toward a time- and position-sensitive detector whose output is analysed to determine the three-dimensional velocity of each incident particle. The PHOTOLOC (photoinitiated reaction analysed using the law of cosines) method[12] is used to map the observed laboratory-frame product speeds $|\mathbf{v}_i|$ into unique centre-of-mass scattering angles θ_i, allowing the measured speed distribution to be converted into the differential cross-section $I(\theta)$.

In Fig. 1 we present the differential cross-sections for the formation of $D_2(v' = 3, j' = 0, 2, 4, 6, 8)$. In all cases a large peak in the forward-scattered direction (the same direction as the initial direction of each collision partner) is observed, although we see that a second, side-scattered peak begins to appear as j' decreases, for $j' \le 4$. This behaviour contradicts conventional wisdom for neutral–neutral inelastic scattering[3]. To seek an explanation, we carry out quasi-classical trajectory (QCT) calculations on the BKMP2 potential energy surface[13]. We note that for the $H + D_2$ reaction at a similar collision energy (2.20 eV), calculations on the BKMP2 surface using QCT and fully quantum mechanical methods have been shown to agree to a high level of accuracy[14].

The QCT methodology has been described in detail previously[15,16]. An overview of the QCT method and the potential energy surface is presented in Supplementary Figs 1–4. Five million $H + D_2(v = 0, j = 0)$ trajectories are propagated at a collision energy of 1.72 eV. The initial and final atom–diatom distances are both 6 Å and the maximum impact parameter is 1.6 Å. Trajectories are analysed on the fly and the number of times the system crosses the reaction barrier is counted (a barrier crossing event is defined as the H–D distance becoming smaller than the D–D bond length, or vice versa). Inelastic trajectories have an even number of crossings, whereas reactive trajectories have an odd number. 2,424,078 trajectories are found

[1]Laser Chemistry, Spectroscopy and Dynamics Group, School of Chemistry, University of Bristol, Bristol BS8 1TS, UK. [2]Department of Chemistry, University of Durham, Durham DH1 3LE, UK. [3]Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA.
*These authors contributed equally to this work.

to be inelastically scattered, and a standard binning methodology[15] gives 842 that produce $D_2(v' = 3, j' = 0)$. Because a significant fraction of inelastic trajectories involve two or more crossings of the reaction barrier, this inelastic scattering process is closely related to reactive scattering. We therefore focus on $D_2(v' = 3, j' = 0)$ instead of j' states with larger cross-sections and purer forward scattering, because of extensive existing work on the complementary reactive channel, $H + D_2 \rightarrow HD(v' = 3, j' = 0) + D$ (refs 16–25).

The reactive channel, like the inelastic channel, had been expected to proceed by means of a back-scattered direct-recoil mechanism, until a surprising forward-scattered peak in the $HD(v' = 3, j' = 0)$ differential cross-section was observed at a collision energy of 1.64 eV (ref. 17). QCT[16,17] and quantum mechanical[18,19] calculations agree that the forward-scattered peak arises from an indirect mechanism that is delayed ~25 fs relative to the direct-recoil mechanism. QCT methods, however, underestimate the intensity of the forward-scattered peak by a factor of roughly two. The reason for this discrepancy is that the forward-scattered features are influenced by interference between near-side and far-side partial waves[23,24]. As shown in Fig. 1, QCT calculations closely agree with experiments for inelastically scattered $D_2(v' = 3, j' = 4, 6, 8)$, whereas the intensity of the forward-scattered feature predicted by QCT is too small for $D_2(v' = 3, j' = 0, 2)$. This finding suggests that a similar interference effect is involved in inelastic scattering with low values of j', and that a fully quantum mechanical treatment is needed to describe the observables of the system completely. Because quantum mechanical calculations for the differential cross-sections of the inelastic channel are not available in the literature, we compare in Supplementary Fig. 7 the quantum mechanical calculations of ref. 19 for reactively scattered $HD(v' = 3, j' = 0)$ with recent experiments using methods identical to those in the present work[25]. The excellent agreement supports the accuracy of both the BKMP2 potential energy surface and the present experiment on the complementary $D_2(v' = 3, j' = 0)$ inelastic scattering channel.

In Fig. 2 we present the correlation between deflection angle and impact parameter for the $D_2(v' = 3, j' = 0)$ trajectories, which agrees with conventional wisdom in that low-impact-parameter collisions are back scattered whereas high-impact-parameter collisions are forward scattered. What is surprising, however, is that high-impact-parameter collisions are effective in forming this highly vibrationally excited product. Furthermore, a correlation is found in which low-impact-parameter trajectories exhibit hard collisions exclusively leading to re-crossing of the barrier for forming HD + D, whereas high-impact-parameter collisions are more glancing in nature and show a reduced amount of re-crossing. Examination of the trajectories shows that the D–D bond is never compressed by these inelastic collisions, irrespective of whether the scattering is forward or backward. On the contrary, the D–D bond is stretched, which results from the pull of the approaching H atom on the closest D atom of D_2. This pull is stronger than would be expected on the basis solely of the van der Waals well depth for the equilibrium D_2 bond length, because the well deepens as the D–D bond lengthens (in a reactive collision, the well deepens enough to capture the H atom and form a new H–D bond).

In those collisions with high impact parameters, which dominate the process because of the weighting with impact parameter, we find that the reactants are scattered in a forward direction. At large impact parameters, the H atom approaches the D_2 diatom in an orientation in which the H atom's motion is nearly perpendicular to the D–D bond axis. The H atom passes through a well for the collinear configuration, but the well does not deepen sufficiently to capture the H atom. Simultaneously the H atom pulls on the nearby D atom, causing vibrational excitation of the D_2 diatom as the H atom escapes the shallow well with diminished kinetic energy. The H atom is not appreciably deflected from its initial direction, causing it to be forward scattered. Figure 3 illustrates this behaviour for one particular high-impact-parameter trajectory, in the form of snapshots of the H–D–D configuration as a function of time. Several representative trajectories are made available in Supplementary Videos 1–6. The lowest impact-parameter collisions forming $D_2(v' = 3, j' = 0)$ are collinear and cross the reaction barrier more than once. Some evidently go on to react and form HD + D, but others are channelled into an outward-going symmetric stretch of the H–D–D complex, which prevents the H atom from being captured by the neighbouring D atom. In this sense, these inelastic scattering events are frustrated reactive scattering events.

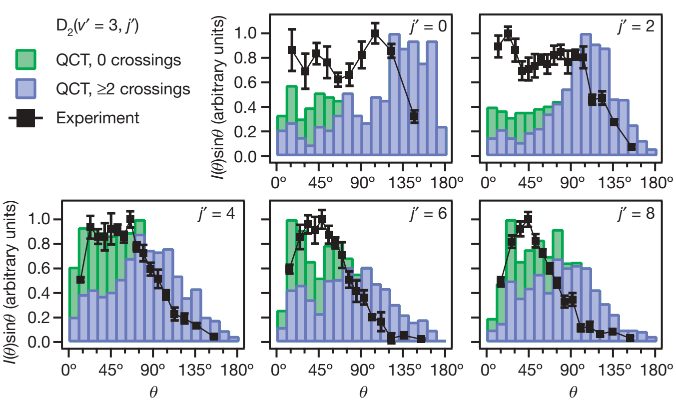


Figure 1 | Products of inelastic $H + D_2$ collisions are mostly forward scattered. Experimental and calculated (QCT, quasi-classical trajectory) differential cross-sections for $H + D_2(v = 0, j = 0, 2) \rightarrow H + D_2(v' = 3, j' = 0, 2, 4, 6, 8)$ inelastic collisions. Each experiment is repeated between three and seven times and the standard deviation of the replicate measurements used as an estimate of the overall error (shown in error bars). Both experiment and theory are scaled so that the peak in the differential cross-section is unity. The scattering angle θ is measured relative to the initial directions of the collision partners, so $\theta = 0^\circ$ represents perfect forward scattering (no deflection) of each collision partner. Inelastic scattering is dominated by re-crossing trajectories.

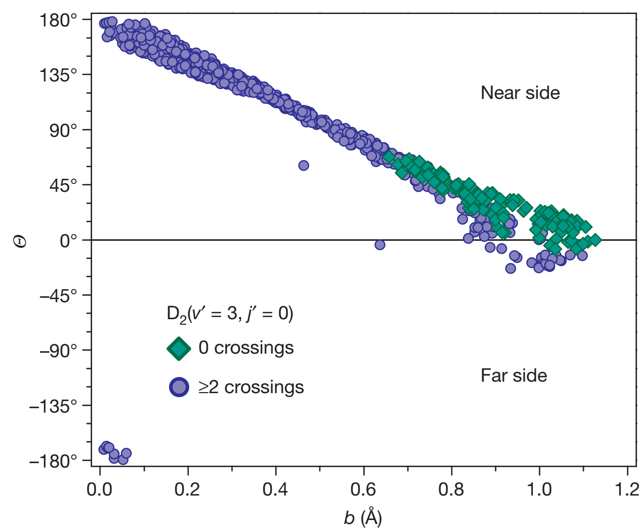


Figure 2 | Impact parameter is linearly correlated with deflection angle. Deflection angle Θ versus impact parameter b, for $D_2(v' = 3, j' = 0)$ inelastically scattered products from QCT calculations of $H + D_2(v = 0, j = 0)$ collisions at a collision energy of 1.72 eV. The strong correlation between impact parameter and scattering direction is clearly shown, with the direct trajectories having high impact parameters and being forward scattered. The deflection angle is defined as the positive scattering angle for near-side scattering (same hemisphere as the incoming H atom) and as the negative scattering angle for far-side scattering. Because of the cylindrical symmetry in the experiment, the laboratory observation is the absolute value of the deflection angle, that is, the scattering angle.

Previous work has shown that QCT overestimates the number of re-crossing trajectories (that is, more trajectories lead to inelastic scattering by crossing back into the $H + D_2$ configuration instead of forming $HD + D$ products) in comparison with fully converged quantum mechanical calculations[22]. As shown in Figs 1 and 2, the majority of the $D_2(v' = 3, j' = 0)$ trajectories re-cross the barrier, particularly at large scattering angles. Therefore, the quantum mechanical differential cross-section would probably show less back scattering than the QCT differential cross-section, in line with the experimental result (Fig. 1). Nevertheless, the QCT calculations provide us with deep insight into the mechanisms of inelastic scattering. Because the QCT calculations and experiments are in close agreement for high-j' levels of $D_2(v' = 3)$, which show one forward-scattered peak, it might be expected that the QCT calculation for $D_2(v' = 3, j' = 4)$, for example, would be more representative of the behaviour of the system. As shown in Supplementary Videos 7–10, the same tug-of-war mechanism is operative for both $D_2(v' = 3, j' = 0)$ and $D_2(v' = 3, j' = 4)$; that is, extension of the D–D bond by the incoming H atom causes vibrational excitation of the D_2 diatom. Again, we emphasise that $D_2(v' = 3, j' = 0)$ has both a forward- and a back-scattered peak. It might be expected that the back-scattered peak could be explained by compression of the D–D bond, but our work shows otherwise.

Forward scattering in vibrational inelastic collisions has been observed previously in ion–molecule reactions, but not to our knowledge in neutral–neutral scattering events. Vibrationally inelastic

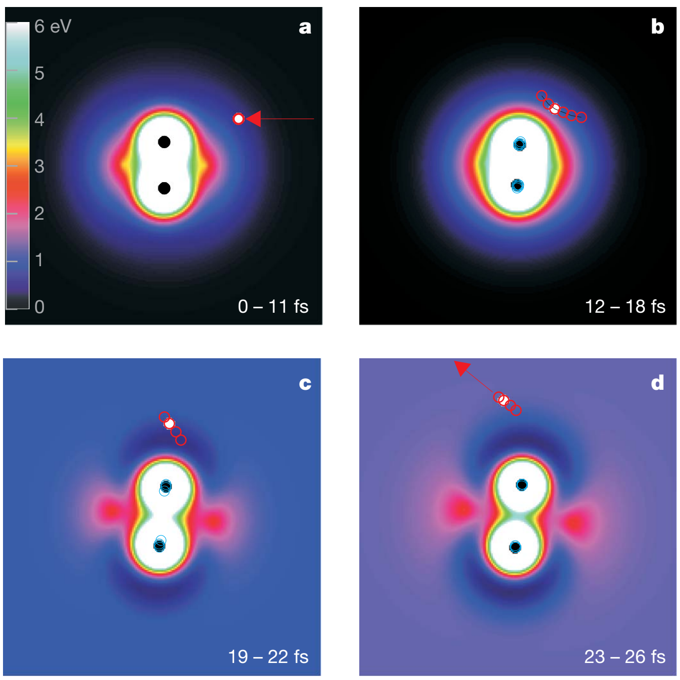


Figure 3 | Snapshots from a representative forward-scattered trajectory. Representative classical trajectory for a $H + D_2(v = 0, j = 0) \rightarrow H + D_2(v' = 3, j' = 0)$ inelastic collision at 1.72 eV, with a scattering angle of 40° (red arrow in **d**) with respect to the initial direction of the H atom (red arrow in **a**). The HD_2 potential energy surface is presented in a space-fixed frame centred at the middle of the D_2 bond. The 0–6-eV energy scale of the potential energy surface is shown in the colour bar: 0 eV corresponds to separation of the H atom and the D_2 diatom. High values of the potential energy are found perpendicular to the D_2 bond, owing to the conical intersection of the HD_2 molecule in the shape of an equilateral triangle (D_{3h} geometry). The red circles represent 1.0-fs time steps of the motion of the H atom and the blue circles show the corresponding motions of the D atoms. The filled circles (white, H atom; black, D atoms) correspond to the step for which the potential energy surface is displayed. The potential energy surfaces are shown for the following D_2 bond lengths: 0.86 Å (11 fs, **a**), 0.72 Å (16 fs, **b**), 0.95 Å (21 fs, **c**) and 1.14 Å (25 fs, **d**). The D_2 bond stretches as the H atom passes through the attractive well and escapes.

scattering of H^+ with H_2, HD and D_2 shows forward scattering[26], which was attributed to 'bond dilution' in which the passing proton withdraws electron density from the diatomic target, thus stretching the bond and inducing vibration. Reference 27 reported a complementary mechanism in the vibrationally inelastic scattering of H^- with H_2, N_2, O_2 and CO_2, which is caused by transient charge transfer into an antibonding orbital of the target. These ion–molecule collisions are fundamentally different from what we report because there the addition or withdrawal of electron density affects the entire molecular geometry, whereas in $H + D_2$ it is only the D atom nearest to the H atom that is affected.

The conventional wisdom that neutral–neutral systems require a sudden compression of the bond must be revised. Hints of this fact were already apparent in ref. 28, which is a study of vibrational relaxation (conversion of vibration to translation and rotation) in radical–radical collisions. There it was concluded that transfer of energy out of vibration is rapidly facilitated as the intramolecular attractive force grows. We believe that this behaviour is part of the same general picture and demonstrates that tug-of-war collisions, a form of frustrated reactive collisions, can be important for open-shell collision systems having strong attractive forces arising from chemical bonding.

Tug-of-war collisions are a new mechanism for inelastic scattering that should be considered whenever it is possible for the collision system to form chemical bonds between the reactants. It must be realized, however, that such wells are not always accessible from the ground states of the collision partners, an example being $H + CO$ (ref. 5). In the case of $H + D_2$, the well depth at the equilibrium D_2 bond length is quite small, but the well deepens rapidly with D–D extension. Collisions with vibrationally excited reactants would also be expected to promote this tug-of-war vibrational inelastic scattering mechanism.

Received 20 December 2007; accepted 2 May 2008.

1. Faubel, M. Vibrational and rotational excitation in molecular collisions. *Adv. At. Mol. Phys.* **19**, 345–394 (1983).
2. Krajnovich, D. J., Parmenter, C. S. & Catlett, D. L. Jr. State-to-state vibrational transfer in atom-molecule collisions. Beams vs. bulbs. *Chem. Rev.* **87**, 237–288 (1987).
3. Levine, R. D. *Molecular Reaction Dynamics* 376 (Cambridge Univ. Press, Cambridge, UK, 2005).
4. Miller, W. H. The semiclassical nature of atomic and molecular collisions. *Acc. Chem. Res.* **4**, 161–167 (1971).
5. McBane, G. C., Kable, S. H., Houston, P. L. & Schatz, G. C. Collisional excitation of CO by 2.3 eV H atoms. *J. Chem. Phys.* **94**, 1141–1149 (1991).
6. Kreutz, T. G. & Flynn, G. W. Analysis of translational, rotational, and vibrational energy transfer in collisions between CO_2 and hot hydrogen atoms: The three-dimensional "breathing" ellipsoid model. *J. Chem. Phys.* **93**, 452–465 (1990).
7. Wight, C. A., Donaldson, D. J. & Leone, S. R. A two-laser pulse-and-probe study of T-R, V energy transfer collisions of H + NO at 0.95 and 2.2 eV. *J. Chem. Phys.* **83**, 660–667 (1985).
8. Rapp, D. & Kassal, T. The theory of vibrational energy transfer between simple molecules in nonreactive collisions. *Chem. Rev.* **69**, 61–102 (1969).
9. Koszinowski, K., Goldberg, N. T., Pomerantz, A. E. & Zare, R. N. Construction and calibration of an instrument for three-dimensional ion imaging. *J. Chem. Phys.* **125**, 133503 (2006).
10. Goldberg, N. T., Koszinowski, K., Pomerantz, A. E. & Zare, R. N. Doppler-free ion imaging of hydrogen molecules produced in bimolecular reactions. *Chem. Phys. Lett.* **433**, 439–443 (2007).
11. Koszinowski, K. *et al.* Differential cross section for the $H+D_2 \rightarrow HD(v' = 1, j' = 2, 6, 10)+D$ reaction as a function of collision energy. *J. Chem. Phys.* **127**, 124315 (2007).
12. Shafer, N. E., Orr-Ewing, A. J., Simpson, W. R., Xu, H. & Zare, R. N. State-to-state differential cross sections from photoinitiated bulb reactions. *Chem. Phys. Lett.* **212**, 155–162 (1993).
13. Boothroyd, A. I., Keogh, W. J., Martin, P. G. & Peterson, M. R. A refined H_3 potential energy surface. *J. Chem. Phys.* **104**, 7139–7152 (1996).
14. Wrede, E. *et al.* The dynamics of the hydrogen exchange reaction at 2.20 eV collision energy: Comparison of experimental and theoretical differential cross sections. *J. Chem. Phys.* **110**, 9971–9981 (1999).
15. Greaves, S. J., Murdock, D., Wrede, E. & Althorpe, S. C. New, unexpected, and dominant mechanisms in the hydrogen exchange reaction. *J. Chem. Phys.* **128**, 164306 (2008).

16. Greaves, S. J., Murdock, D. & Wrede, E. A quasi-classical trajectory study of the time-delayed forward scattering in the hydrogen exchange reaction. *J. Chem. Phys.* **128,** 164307 (2008).
17. Fernández-Alonso, F. *et al.* Evidence for scattering resonances in the $H+D_2$ reaction. *Angew. Chem. Int. Ed.* **39,** 2748–2752 (2000).
18. Allison, T. C., Friedman, R. S., Kaufman, D. J. & Truhlar, D. G. Analysis of the resonance in $H+D_2 \rightarrow HD(v'=3)+D$. *Chem. Phys. Lett.* **327,** 439–445 (2000).
19. Althorpe, S. C. *et al.* Observation and interpretation of a time-delayed mechanism in the hydrogen exchange reaction. *Nature* **416,** 67–70 (2002).
20. Fernández-Alonso, F. & Zare, R. N. Scattering resonances in the simplest chemical reaction. *Annu. Rev. Phys. Chem.* **53,** 67–99 (2002).
21. Ayers, J. D. *et al.* Measurement of the cross section for $H+D_2 \rightarrow HD(v'=3, j'=0)+D$ as a function of angle and energy. *J. Chem. Phys.* **119,** 4662–4670 (2003).
22. Aoiz, F. J., Bañares, L. & Herrero, V. J. The $H + H_2$ reactive system. Progress in the study of the dynamics of the simplest reaction. *Int. Rev. Phys. Chem.* **24,** 119–190 (2005).
23. Monks, P. D. D., Connor, J. N. L. & Althorpe, S. C. Theory of time-dependent reactive scattering: Cumulative time-evolving differential cross sections and nearside-farside analyses of time-dependent scattering amplitudes for the $H+D_2 \rightarrow HD+D$ reaction. *J. Phys. Chem. A* **110,** 741–748 (2006).
24. Monks, P. D. D., Connor, J. N. L. & Althorpe, S. C. Nearside-farside and local angular momentum analyses of time-independent scattering amplitudes for the $H+D_2(v_i=0, j_i=0) \rightarrow HD(v_f=3, j_f=0)+D$ reaction. *J. Phys. Chem. A* **111,** 10302–10312 (2007).
25. Goldberg, N. T., Zhang, J., Miller, D. J. & Zare, R. N. Corroboration of theory for $H+D_2 \rightarrow D+HD(v'=3, j'=0)$ reactive scattering dynamics. *J. Phys. Chem. A*. doi: 10.1021/jp801187p (in the press).
26. Giese, C. F. & Gentry, W. R. Classical trajectory treatment of inelastic scattering in collisions of H^+ with H_2, HD, and D_2. *Phys. Rev. A* **10,** 2156–2173 (1974).
27. Hege, U. & Linder, F. Vibrationally inelastic scattering of H^- ions from H_2, N_2, O_2 and CO_2. *Z. Phys. A* **320,** 95–104 (1985).
28. Osborn, M. K. & Smith, I. W. M. A quasiclassical trajectory study of vibrational energy transfer in collisions involving intermolecular attraction of moderate strength. *Chem. Phys.* **91,** 13–26 (1984).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The Stanford team gratefully acknowledges support by the US National Science Foundation under grant NSF CHE 0650414. S.J.G. is supported by the EPSRC LASER Portfolio Partnership grant GR/S71750/01.

Author Contributions N.T.G. designed and performed the experiments with the help of J.Z. and D.J.M., analysed the experimental data, and assisted with the manuscript. S.J.G and E.W. performed quasi-classical trajectory calculations, formulated the mechanism, and prepared figures and movies. R.N.Z. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.N.Z. (zare@stanford.edu).

Vol 454 | 3 July 2008 | doi:10.1038/nature07102

LETTERS

A light carbon reservoir recorded in zircon-hosted diamond from the Jack Hills

Alexander A. Nemchin[1], Martin J. Whitehouse[2], Martina Menneken[3], Thorsten Geisler[3], Robert T. Pidgeon[1] & Simon A. Wilde[1]

The recent discovery of diamond–graphite inclusions in the Earth's oldest zircon grains (formed up to 4,252 Myr ago) from the Jack Hills metasediments in Western Australia[1] provides a unique opportunity to investigate Earth's earliest known carbon reservoir. Here we report ion microprobe analyses of the carbon isotope composition of these diamond–graphite inclusions. The observed $\delta^{13}C_{PDB}$ values (expressed using the PeeDee Belemnite standard) range between −5 per mil and −58 per mil with a median of −31 per mil. This extends beyond typical mantle values of around −6 per mil to values observed in metamorphic and some eclogitic diamonds that are interpreted to reflect deep subduction of low-$\delta^{13}C_{PDB}$ biogenic surface carbon. Low $\delta^{13}C_{PDB}$ values may also be produced by inorganic chemical reactions[2], and therefore are not unambiguous evidence for life on Earth as early as 4,250 Myr ago. Regardless, our results suggest that a low-$\delta^{13}C_{PDB}$ reservoir may have existed on the early Earth.

Detrital zircon ($ZrSiO_4$) grains from the ~3,000-Myr-old Jack Hills metaconglomerate in Western Australia range in age up to ~4,400 Myr and currently represent the only direct samples of the Hadean Earth (earlier than ~4,000 Myr ago). Intense study of these zircons, primarily by microanalytical methods, has yielded an abundance of invaluable data relevant to models of the Earth's earliest tectonic regime. Nonetheless, models for the Hadean eon still range from the uniformitarian, characterized by the generation of voluminous continental crust by subduction-dominated tectonics[3] analogous to modern tectonic processes that are thought to have operated since ~3,900 Myr ago[4], to a radically different picture of mafic-dominated, potentially long-lived crust, evidence for which largely comes from accessible geochemical and isotopic reservoirs sampled since the Hadean[5].

The recent discovery of diamond and graphite inclusions in 3,050–4,250-Myr-old zircon grains from the Jack Hills[1] introduces a new dimension to the debate on the early evolution of the Earth. Most profound is the indication of a carbon reservoir and processes capable of crystallizing diamond and graphite. The origin of the carbon and the nature of the carbon reservoir, as well as the process by which microdiamonds can be incorporated in zircon together with 'granitic' inclusions[6], present problems fundamental to understanding processes active in the early history of the Earth. Carbon isotopes in the diamond and graphite inclusions provide a basic parameter critical for resolving these problems. We report the composition of 22 diamond and graphite inclusions from 18 Jack Hills zircon grains determined by ion microprobe (see Methods).

Diamond and graphite can readily be distinguished by Raman spectroscopy and their contrasting cathodoluminescence intensities (Supplementary Figs 1–3), diamond inclusions being cathodoluminescence bright in comparison with cathodoluminescence-dark graphite. The inclusions are classified (Table 1), on the basis of Raman and cathodoluminescence response, as being either exclusively diamond (including some polycrystalline aggregates of nanometre- to micrometre-sized diamond) or mixtures of diamond and graphite. Even in inclusions classified as being mainly graphite, the graphite fraction may actually be low, because the strong light absorption of a graphite surface layer only a few nanometres thick would strongly attenuate cathodoluminescence and Raman signals.

Inclusions identified as being mainly graphite exhibit a range of $\delta^{13}C_{PDB}$ values between about −51‰ and −15‰, whereas $\delta^{13}C_{PDB}$ values in most diamond inclusions (except JH3-189-1) range between about −24‰ and −5‰ (Table 1), suggesting a difference in the carbon isotopic composition of diamond and graphite inclusions (Fig. 1). However, diamond inclusion JH3-189-1 does not follow this trend, as it has the lowest $\delta^{13}C_{PDB}$ values (−50.9 ± 1.3‰ and −58.0 ± 5.2‰; Table 1). The results of comprehensive tests of potential analytical problems discount the possibility that the observed low $\delta^{13}C_{PDB}$ values are analytical artefacts (see Supplementary Information). We further note that it is highly unlikely that an apparent difference of about 20‰ between diamond and graphite is the result of a solid-state phase transition. Consequently, it is not evident on the basis of available data that there is a systematic difference in carbon isotope composition between graphite- and diamond-dominated inclusions. The apparent trend could simply be the result of a statistically small population of analysed inclusions and must be tested in further studies of carbon inclusions in Jack Hills zircons.

There is also no significant correlation between $\delta^{13}C_{PDB}$, diamond size, and the age of the host zircon grains (Table 1, Fig. 1). Some zircon grains have more than one diamond–graphite inclusion, and a comparison of the carbon isotope compositions of these inclusions shows no systematic trends. Some grains contain inclusions with similar carbon isotope compositions (for example, two inclusions in zircon JH3-48 have $\delta^{13}C_{PDB}$ values of −34.2 ± 3.1‰ and −34.3 ± 1.9‰), whereas other zircons have inclusions with very different $\delta^{13}C_{PDB}$ values (for example, two inclusions in the oldest zircon grain, JH3-20, with an age of 4,252 ± 7 Myr (ref. 1), show the maximum difference in $\delta^{13}C_{PDB}$ values (Table 1)).

Previously[1] we suggested that all diamond inclusions may be older than 4,250 Myr, based on hafnium isotope studies of the Jack Hills zircon population indicating that younger zircon grains may have formed by repeated recycling of a >4,200-Myr-old source[7]. However, regardless of the possibility of several episodes of diamond formation, any given inclusion must be at least as old as its host zircon. The observed large variation in $\delta^{13}C_{PDB}$ values in inclusions hosted in the same zircon grain suggests that the carbon inclusions formed from different material and/or under different geological conditions before they were eventually included in the zircon. We

[1]Department of Applied Geology, Western Australian School of Mines, Curtin University of Technology, Bentley, Western Australia 6102, Australia. [2]Laboratory for Isotope Geology, Swedish Museum of Natural History, SE-104 05 Stockholm, Sweden. [3]Institut für Mineralogie, Westfälische Wilhelms-Universität, Corrensstrasse 24, 48149 Münster, Germany.

Table 1 | Analysis of the inclusions

Label	Type†	Zircon growth texture‡	207Pb/206Pb age (Myr)§	Detection mode‖	$\delta^{13}C_{PDB}$ (‰)
JH15-497-1	dia	oz	3,455±10	7/FC	−18.5±1.0
				7/FC	−13.3±1.2
				2/EM	−18.0±2.6
JH3-20-3	dia	oz	4,252±14	7/FC	−11.8±0.8
				2/EM	−4.8±2.9
				2/EM	−4.9±2.6
				2/EM	−16.5±3.2
JH3-50-4	dia	iz	3,681±10	7/FC	−19.3±1.3
JH11-4-1	dia	oz	3,370±24	7/FC	−24.2±1.0
JH11-250-1	dia*	pz	3,561±16	7/FC	−20.2±5.1
JH15-444-1	dia	oz	3,346±8	7/FC	−12.5±1.9
				7/FC	−17.4±1.5
				2/EM	−17.0±4.7
JH3-189-1	dia*	sz	3,369±12	7/FC	−50.9±1.3
				2/EM	−58.0±5.2
JH3-48-1	dia/gr	cbz	3,454±14	7/FC	−34.2±3.1
JH15-243-1	dia/gr	pz	3,382±12	7/FC	−41.0±5.9
JH15-142-1	dia/gr	rz	3,897±6	7/FC	−33.6±1.6
				2/EM	−37.8±3.0
JH15-82-2	dia/gr	oz	3,225±20	7/FC	−28.7±1.4
				2/EM	−31.8±4.2
JH15-82-3	dia/gr	oz	3,225±20	7/FC	−30.5±1.3
				2/EM	−29.4±2.7
JH11-61-16	dia*/gr	sz	3,451±14	7/FC	−23.4±0.9
JH11-267-1	dia/gr	oz	3,377±10	7/FC	−31.0±0.8
JH11-61-15	gr	sz	3,451±14	7/FC	−43.1±1.5
JH15-2-3	gr	oz	3,476±10	7/FC	−36.1±6.4
				2/EM	−35.9±4.5
JH11-13-1	gr	sz	3,600±12	7/FC	−40.8±2.0
JH3-165-1	gr	oz	3,266±14	7/FC	−35.6±1.7
JH3-20-1	gr	oz	4,252±14	7/FC	−51.1±1.7
				2/EM	−40.2±8.5
JH3-48-5	gr	iz	3,454±14	7/FC	−34.3±1.9
JH3-124-3	gr	oz	3,058±14	7/FC	−37.9±0.7
				2/EM	−40.3±2.6
				2/EM	−44.0±3.0
				2/EM	−42.3±2.6
				2/EM	−42.1±3.1
				2/EM	−47.1±2.9
JH3-134-4	gr	oz	3,312±20	7/FC	−28.8±1.6
				2/EM	−24.0±3.5
				2/EM	−30.6±3.7
				2/EM	−26.2±3.2
				2/EM	−15.4±3.5
				2/EM	−27.9±3.7

Mineralogy of the analysed inclusions, textural features of the host zircons, $^{207}Pb/^{206}Pb$ ages of the inclusion-bearing zircon growth zone, analytical detection mode and $\delta^{13}C_{PDB}$ values. $\delta^{13}C_{PDB} = \delta^{13}C_{measured} + \delta^{13}C_{PDB(standard)} + 10^{-3}\delta^{13}C_{measured}\delta^{13}C_{PDB(standard)}$, where $\delta^{13}C_{measured} = ((^{13}C/^{12}C)_{sample}/(^{13}C/^{12}C)_{standard} - 1) \times 10^3$.

† dia, mainly diamond; dia/gr, diamond and graphite; gr, mainly graphite; dia*, micrometre-sized, polycrystalline diamond detected by Raman spectroscopy[1].

‡ oz, oscillatory zoning; iz, irregular zoning; sz, sector zoning; pz, planar zoning; cbz, cross-bedded zoning; rz, re-equilibration zone (see Supplementary Figs 1–3). Late overprint resulting in the formation of homogeneously grey patches in the zircon is not considered in this classification.

§ Data from ref. 1.

‖ 7/FC: Multicollector analyses with a 7-μm spot size; 2/EM: Mono-collector analyses using an electron multiplier and a 2-μm spot size.

consider it unlikely that the diamonds post-date their host zircon, perhaps formed from a supercritical COH-rich fluid that infiltrated the host zircon grains along the cracks that are visible in most diamond-bearing zircon grains (Supplementary Figs 1–3), as such a fluid would have had a homogeneous carbon isotope composition inconsistent with the extreme variation in $\delta^{13}C_{PDB}$ values observed in different inclusions hosted by the same zircon grain.

In addition, a supercritical COH-rich fluid would have been highly reactive and not in equilibrium with any of the Jack Hills zircons. Thus, such a fluid would probably have reacted with the zircon to form re-equilibration textures[8] or baddeleyite (monoclinic ZrO_2)[9]. However, no indications of any type of reaction along the cracks have been observed in any of the Jack Hills zircons. We further emphasize that the carbon inclusions occur both in central and outer parts of some host grains, as well as in texturally different parts of complexly zoned zircon (Table 1, Supplementary Figs 1–3), suggesting that their inclusion into the zircon is not related to any specific process that formed the distinct growth or re-equilibration zones. Some inclusions are located within irregularly zoned parts of zircon grains that have not always been directly dated, leaving open the possibility that these inclusions could be related to processes younger than the dated parts of the grains. Alternatively, these parts can represent zircon with an irregular trace element content, which is a common feature of zircon growth around a foreign object[10].

It has been suggested that carbon was included in zircons of different ages as graphite and then transformed into diamond during a single, high-pressure metamorphic event that occurred less than 3,000 Myr ago[11]. However, this contradicts the low compressibility of zircon (for example only a ~1.6% density increase at room temperature and 4 GPa)[12] when compared with the very large density increase accompanying the first-order graphite-to-diamond transformation (for example ~44% at room temperature and 4 GPa)[13], which precludes a graphite inclusion in zircon being transformed into diamond. Therefore, the simplest explanation, and the one which is supported by most observations, is that the diamond formation must pre-date zircon crystallization and, most probably, is not related to zircon formation.

The fact that at least some inclusions are probably as old as 4,250 Myr provides us with the oldest terrestrial carbon reservoir discovered so far[1]. We interpret the range of $\delta^{13}C_{PDB}$ values observed in these inclusions as a true measure of a potentially heterogeneous early-Earth carbon reservoir. Other than those in the Jack Hills, the oldest terrestrial diamonds found are associated with the kimberlite–lamproite suite of rocks[13]. These diamonds are of peridotite type (P-diamonds), on the basis of their inclusion paragenesis. They are about 3,000 Myr old, or slightly older, and have $\delta^{13}C_{PDB}$ values indistinguishable from that of the bulk Earth[14]. Light carbon isotope compositions are commonly found in so-called eclogitic diamonds (E-diamonds) showing negative $\delta^{13}C_{PDB}$ values (from −0‰ to −40‰), microdiamonds formed in subducted ultrahigh-pressure metamorphic rocks ($\delta^{13}C_{PDB}$ from −3‰ to −30‰), impact-related diamonds ($\delta^{13}C_{PDB}$ from −8‰ to −22‰) and carbonados ($\delta^{13}C_{PDB}$ from −25‰ to −32‰)[15,16]. Additionally, strongly negative $\delta^{13}C_{PDB}$ values (from −15‰ to −40‰) have been observed in sedimentary carbon[15].

These negative values are often taken as indicators of a contribution from a biogenic source which, if applicable to the Jack Hills inclusions, suggests that life might first have appeared on Earth well before the late heavy-meteorite bombardment about 3,900 Myr ago[17]. The significance of low $\delta^{13}C_{PDB}$ values as a unique biomarker remains controversial[2], however, particularly when applied to Paleoarchaean and Eoarchaean rocks from which other evidence for biogenic activity is lacking[18]. Experimental work conducted to investigate carbon isotope fractionation effects associated with abiotic organic synthesis[19,20] has demonstrated that negative $\delta^{13}C_{PDB}$ values similar to those observed in the Jack Hills inclusions can also be produced in reactions involving carbon oxides, methane, hydrogen and water, all significant components of both early-atmosphere and volcanic gases.

By analogy with the interpretation of light carbon in <1,500-Myr-old eclogitic and <500-Myr-old metamorphic diamonds[4], the low $\delta^{13}C_{PDB}$ values of >3,000-Myr-old Jack Hills diamond–graphite inclusions, in combination with the observation that their mineralogical characteristics resemble those of modern ultrahigh-pressure metamorphic diamonds[1], could further be interpreted as evidence for an early Archaean and Hadean tectonic regime that was very similar to present day plate tectonics. In contrast, analysis of the consequences of accretion and the giant moon-forming impact[21,22], as well as investigation of accessible geochemical and isotopic reservoirs, has been used to conclude that the early history of the Earth was characterized by long-lived mafic to ultramafic crust[4] and a mantle thermal regime that prevented significant convection[5]. Such a long-lived crust

could accumulate light carbon as a result of interaction with an early-Earth atmosphere that was rich in methane and possibly carbon dioxide (ref. 5). As mantle convection gradually restarted, this crust could have been recycled back into the mantle, introducing light carbon that over time homogenized with mantle.

Possible explanations for very negative $\delta^{13}C_{PDB}$ values in mantle-derived materials such as diamonds are not necessarily limited to terrestrial sources. Chondrites have bulk $\delta^{13}C_{PDB}$ values between 0‰ and −28‰ (ref. 23), presolar diamonds in primitive carbonaceous chondrites have values as low as −38‰ (refs 24, 25), and extremely low values of $\delta^{13}C_{PDB}$ (the majority falling between −20‰ and −50‰) have been reported from some intraplanetary dust particles[26]. This has led to the suggestion that the variation of carbon isotope composition observed in mantle-derived materials is a primordial feature inherited from the time of accretion[14].

The observations from the Jack Hills carbon inclusions of a $\delta^{13}C_{PDB}$ range of about 50‰, with a number extending beyond the lower limits (about −38‰ for the eclogite diamonds[15]) established for all other diamonds, clearly distinguishes them from peridotitic diamonds and carbonados, raising the possibility that these features are unique for the early Earth. Even most biogenic carbon[27] is heavier than many of the carbon samples from the Jack Hills inclusions, although the organic carbon record includes two $\delta^{13}C_{PDB}$ minima, recorded in the Fortescue group (Australia) and in the Francevillian formation of Gabon (Africa), which extend down to values of –50‰. These minima have been interpreted as evidence of methanogenic carbon fixation[27]. Mixing of such carbon with mantle-derived carbon may produce the observed variation in $\delta^{13}C_{PDB}$. The range of $\delta^{13}C_{PDB}$ values observed in carbonaceous chondrites is also different from that of the Jack Hills inclusions. Therefore, if Jack Hills diamonds were formed by impacts (although there is Raman spectroscopic evidence that this is not the case[1]), at least some of the carbon required for the formation of these diamonds must have come from impacted terrestrial rocks, where it would still have required a process capable of producing very negative $\delta^{13}C_{PDB}$ values. The only extraterrestrial source of such light carbon is some intraplanetary dust particles and/or presolar grains[26]. If these were indeed the source of light carbon in the Jack Hills inclusions, a substantial flux would have been required.

All these possible sources of carbon involve accumulation of light carbon near the surface and require a mechanism either to transfer it into the diamond stability field (unless the diamond grew outside its stability field[14]) or to transfer impact-produced diamonds at least to the region where zircon is forming. Possible transfer mechanisms include modern-style plate tectonic subduction and vertical tectonics ('sagduction')[28], a process that might have operated in the Archaean. Some inorganic reactions (for example the Fischer–Tropsch process) may produce both diamond and the observed spread in $\delta^{13}C_{PDB}$ values[2] at a variety of depths and conditions, and might obviate the need for additional transfer mechanisms.

We are aware that many of our observations remain unexplained and that conclusions based on them are thus controversial. Specifically, the observed, strongly negative $\delta^{13}C_{PDB}$ values in the Jack Hills diamond–graphite inclusions do not indicate a unique source of carbon. We caution that they should not be interpreted as prima-facie evidence for biological activity in the Hadean, although they do not exclude such a possibility. Whatever the process responsible for the formation of light carbon observed in the Jack Hills zircon-hosted inclusions, they most likely predate their respective host grains. Therefore, our data suggest that a light carbon reservoir may have existed from the earliest stages of Earth's history. It is thus conceivable that the low $\delta^{13}C_{PDB}$ signature exhibited by many crust- and mantle-derived rocks throughout the Earth's history may in part reflect recycling of this reservoir.

METHODS SUMMARY

Carbon isotope compositions of diamond graphite inclusions were analysed by secondary-ion mass spectrometry using a Cameca IMS1270 instrument (Nordsim facility, Stockholm) operating with instrument parameters typical for light stable isotope analysis (see, for example, ref. 29). Analyses were performed using two different spatial resolutions and secondary-ion detection modes. The first used a nominal 2-μm spot with ^{12}C and ^{13}C signals measured in peak jumping mode on an ion-counting electron multiplier (electronically gated dead time of 44 ns); the second used a nominal 7-μm spot with simultaneous measurement of ^{12}C and ^{13}C on Faraday detectors. In order to eliminate interference of the $^{12}C^{1}H^{+}$ species on $^{13}C^{+}$, all analyses were performed at a nominal mass resolution ($M/\Delta M$) of 4,800. Potential sample-charging effects due to the small size of the inclusions embedded in a non-conducting matrix were minimised using a low-energy, normal-incidence electron flood gun.

Data are reported as $\delta^{13}C_{PDB}$ values in parts per thousand (‰), where PDB refers to the carbon isotope composition of the PeeDee Belemnite. A synthetic diamond, SYNAL, with $\delta^{13}C_{PDB} = -23.9$‰ (ref. 30) was used as a reference material. This was mounted in epoxy, coated with gold, and inserted alongside the Jack Hills zircon mounts in the same sample holder to enable bracketing of sample and standard analyses. Each analysis consisted of a 120-s pre-sputter period to remove the gold coating from an area larger than the analytical spot, followed by automated centring of the beam in the field aperture and centring of the magnet field. Multicollector analyses consisted of 16 cycles of 10-s integrations. Mono-collector analyses consisted of 50 cycles through the masses with integration times of 2 s and 6 s on the ^{12}C and ^{13}C peaks, respectively.

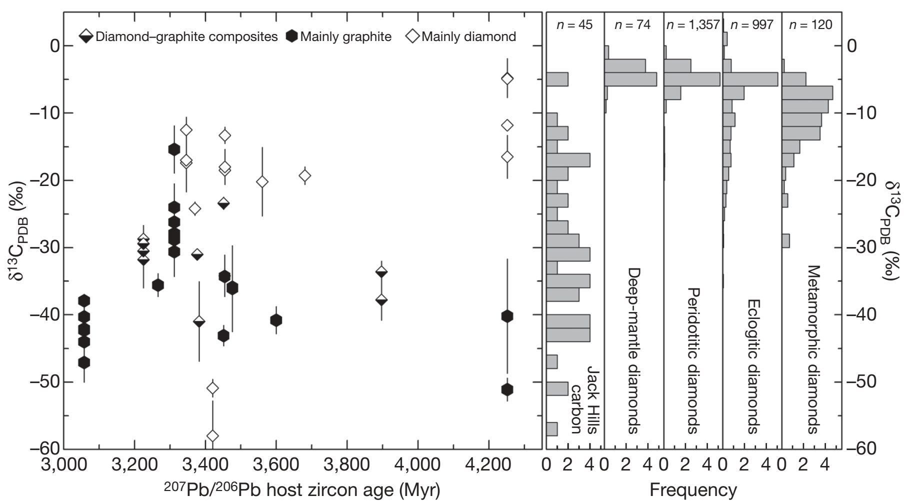


Figure 1 | $\delta^{13}C_{PDB}$ values of diamond–graphite inclusions. We plot the values against the age of the Jack Hills host zircons, and provide a comparison of their distribution with $\delta^{13}C_{PDB}$ distributions of diamonds from known geological settings. The $\delta^{13}C_{PDB}$ ranges of known diamond sources were redrawn from ref. 15. Error bars, 2 s.d.; *n*, number of analyses.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 11 September 2007; accepted 14 May 2008.

1. Menneken, M., Nemchin, A. A., Geisler, T., Pidgeon, R. T. & Wilde, S. A. Hadean diamonds in zircon from Jack Hills, Western Australia. *Nature* **448,** 917–920 (2007).
2. Horita, J. Some perspectives on isotope biosignatures for early life. *Chem. Geol.* **218,** 171–196 (2005).
3. Harrison, T. M. *et al.* Heterogeneous Hadean hafnium: evidence of continental crust at 4.4 to 4.5 Ga. *Science* **310,** 1947–1950 (2005).
4. Shirey, S. B., Kamber, B. S., Whitehouse, M. J., Mueller, P. A. & Basu, A. R. in *When Did Plate Tectonics Start on Earth?* (eds Condie, K. C. & Pease, V.) 1–29 (Special Paper 440, Geological Society of America, Boulder, Colorado, 2008).
5. Kramers, J. D. Hierarchical Earth accretion and the Hadean eon. *J. Geol. Soc. Lond.* **164,** 3–17 (2007).
6. Maas, R., Kinny, P. D., Williams, I. S., Froude, D. O. & Compston, W. The Earth's oldest known crust: A geochronological and geochemical study of 2900–4200 Ma old zircons from Mt Narryer and Jack Hills, Western Australia. *Geochim. Cosmochim. Acta* **56,** 1281–1300 (1992).
7. Amelin, Y., Lee, D.-C. & Halliday, A. N. Early–middle Archaean crustal evolution deduced from Lu–Hf and U–Pb isotopic studies of single zircon grains. *Geochim. Cosmochim. Acta* **64,** 4205–4225 (2000).
8. Geisler, T., Schaltegger, U. & Tomaschek, F. Re-equilibration of zircon in aqueous fluids and melts. *Elements* **3,** 45–51 (2007).
9. Rizvanova, N. G. *et al.* Zircon reaction and stability of the U-Pb isotope system during interaction with carbonate fluid: experimental hydrothermal study. *Contrib. Mineral. Petrol.* **139,** 101–114 (2000).
10. Hoskin, P. W. O. & Schaltegger, U. The composition of zircon and igneous and metamorphic petrogenesis. *Rev. Mineral. Geochem.* **53,** 27–62 (2003).
11. Williams, I. S. Old diamonds and the upper crust. *Nature* **448,** 880–881 (2007).
12. Hazen, R. M. & Finger, L. W. Crystal structure and compressibility of zircon at high pressure. *Am. Mineral.* **64,** 196–201 (1979).
13. Kerley, G. I. & Chhabildas, L. *Multicomponent-Multiphase Equation of State for Carbon.* Sandia Report SAND2001–2619 (Sandia National Laboratories, Albuquerque, 2001).
14. Haggerty, S. E. A diamond trilogy: superplumes, supercontinents, and supernovae. *Science* **285,** 851–860 (1999).
15. Cartigny, P. Stable isotopes and the origin of diamonds. *Elements* **1,** 79–84 (2005).
16. Heaney, P. J., Vicenzi, E. P. & De, S. Strange diamonds: The mysterious origins of carbonado and framesite. *Elements* **1,** 85–89 (2005).
17. Tera, F., Papanastassiou, D. A. & Wasserburg, G. J. Isotopic evidence for a terminal lunar cataclysm. *Earth Planet. Sci. Lett.* **22,** 1–21 (1974).
18. Whitehouse, M. J. & Fedo, C. M. in *Earth's Oldest Rocks. Developments in Precambrian Geology* Vol. 15 (eds Van Kranendonk, M. J., Smithies, R. H. & Bennett, V.) 841–853 (Elsevier, Amsterdam, 2007).
19. Horita, J. & Berndt, M. E. Abiogenic methane formation and isotopic fractionation under hydrothermal conditions. *Science* **285,** 1055–1057 (1999).
20. McCollom, T. M. & Seewald, J. S. Carbon isotope composition of organic compounds produced by abiotic synthesis under hydrothermal conditions. *Earth Planet. Sci. Lett.* **243,** 74–84 (2006).
21. Benz, W., Cameron, A. G. W. & Melosh, H. J. The origin of the Moon and the single-impact hypothesis. *Icarus* **81,** 113–131 (1989).
22. Canup, R. N. & Asphaug, E. Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* **412,** 708–712 (2001).
23. Krot, A. N., Keil, K., Goodrich, C. A. & Scott, E. R. D. in *Treatise on Geochemistry, Meteorites, Comets, and Planets* Vol. 1 (ed. Davis, A. M.) 83–128 (Elsevier, Amsterdam, 2003).
24. Russell, S. S., Pillinger, C. T. & Arden, J. W. Evidence for multiple sources of diamond from primitive chondrites. *Science* **254,** 1188–1191 (1991).
25. Verchovsky, A. B. *et al.* C, N, and noble gas isotopes in grain size separates of presolar diamonds from Efremovka. *Science* **21,** 1165–1168 (1998).
26. McKeegan, K. D. *et al.* Isotopic compositions of cometary matter returned by stardust. *Science* **314,** 1724–1728 (2007).
27. Schidlowski, M. A 3,800-million-year isotopic record of life from carbon in sedimentary rocks. *Nature* **333,** 313–318 (1983).
28. Bedard, J. H. A cathalitic delamination-driven model for coupled genesis of Archaean crust and sub-continental lithospheric mantle. *Geochim. Cosmochim. Acta* **70,** 1188–1214 (2006).
29. Nemchin, A. A., Pidgeon, R. T. & Whitehouse, M. J. Re-evaluation of the origin and evolution of >4.2 Ga zircons from the Jack Hills metasedimentary rocks. *Earth Planet. Sci. Lett.* **244,** 218–233 (2006).
30. Harte, B., Fitzsimons, I. C. W., Harris, J. W. & Otter, M. L. Carbon isotope ratios and nitrogen abundances in relation to cathodoluminescence characteristics for some diamonds from the Kaapvaal Province, S. Africa. *Mineral. Mag.* **63,** 829–856 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This project was supported by the Curtin University internal research grant to A.A.N. We thank I. Fitzsimons and P. Kinny for their comments on the earlier version of the manuscript. SYNAL diamond used as a reference in this study was kindly provided by J. Craven of the Edinburgh University Ion Microprobe Facility. The Nordsim facility is operated and funded under an agreement by the joint Nordic research councils (NOS-N); this is Nordsim contribution 200.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.A.N. (nemchina@kalg.curtin.edu.au).

METHODS

Test of diamond–graphite instrumental mass fractionation differences. The complex mineralogical composition of the inclusions, which are mainly represented by mixtures of diamond and graphite in different proportions, raises the possibility of bias in isotope ratios due to different instrumental mass fractionation from the two phases. Although close similarity of instrumental mass fractionation from diamond and graphite has been demonstrated previously[31], these experiments used extreme energy-filtering methods on a small ion microprobe. Under the analytical conditions used for the present study, which specifically did not utilize energy filtering, we investigated the possibility of matrix influence on the observed carbon isotope compositions, by analysing SYNAL diamond against two pyrolysed graphite samples (pyr-1 and pyr-2). Pyr-1 has a conventionally determined $\delta^{13}C_{PDB}$ value of −30.8‰ (G. Layne, personal communication), which was determined using several small subsamples of pyr-1 disk in the Laboratory for Stable Isotope Science at The University of Western Ontario. This value was indistinguishable from the results of a direct $\delta^{13}C_{PDB}$ determination using secondary-ion mass spectrometry at the Woods Hole Oceanographic Institution, made by comparing pyr-1 to a pressed pellet of NBS-24 graphite, which yielded an average $\delta^{13}C_{PDB}$ value of −30.7‰

The results are shown in Supplementary Fig. 4 and Supplementary Table 1 and suggest a $\delta^{13}C_{PDB}$ value of -36.7 ± 0.5‰ (with 95% confidence) for the pyr-2 sample when analysed against the SYNAL diamond and a $\delta^{13}C_{PDB}$ value of -35.7 ± 0.3‰ (with 95% confidence) when analysed against pyr-1 graphite. The errors do not include uncertainties in the 'true' values determined for SYNAL diamond and pyr-1 graphite. The two values are barely outside the errors, suggesting that even if there is a minor matrix effect during the carbon isotope analysis of graphite and diamond, it does not exceed 1‰ and definitely cannot be responsible for the variation in $\delta^{13}C_{PDB}$ of about 50‰ observed in the inclusions.

Test of relief effects on instrumental mass fractionation. A second analytical problem is presented by the difficulty of polishing the inclusions well, as a result of significant difference in hardness between the diamond and graphite in the inclusions and between these two minerals and the host zircon. As a consequence, it is impossible to produce a flat, level surface near the inclusions, which raises a concern that some of the isotopic effects observed during the analysis of unknowns may be influenced by changes in the ion optical configuration near the analysed spots, resulting in different sputtering behaviour of ^{13}C and ^{12}C. Although the problem cannot be eliminated completely it was closely monitored during and after the analyses.

Instrumental mass fractionation associated with surface imperfections during an ion probe analysis is commonly accompanied by a decrease in the counting rate of isotopes. Although about 50% of our analyses of inclusions showed an ^{12}C intensity less than 30% of that observed in the standard, and about 35% of the inclusions were found to produce ^{12}C counting rates greater than half of that in the standard, the observed counting rate and $\delta^{13}C_{PDB}$ values are not correlated (Supplementary Table 2, Supplementary Fig. 5). Two inclusions that show ^{12}C counting rates similar to the standard have respective $\delta^{13}C_{PDB}$ values of -40.3 ± 2.6‰ and -4.9 ± 2.6‰. The multicollector measurements, made using the larger (7-μm) spot, generally show smaller counting rates relative to the standard when compared to the signals measured in peak-jumping mono-collection mode, where a smaller (2-μm) spot was used. This systematic difference suggests that the observed decrease in the ^{12}C counting rates during analysis of inclusions is related to the possible overlap of zircon surrounding inclusions by the analytical spot. This overlap is less likely to occur with the smaller spot size used during the mono-collector analysis. The fact that there is no difference visible, within the errors, between the $\delta^{13}C_{PDB}$ values obtained using large and small spot sizes (Supplementary Fig. 6) suggests that this partial overlap, if it occurs, does not introduce a significant fractionation between ^{12}C and ^{13}C. It also indicates that there is no major fractionation of isotopes caused by the possible relief within the inclusions, as the analyses with the small spot size are expected to be more sensitive to this relief variation and therefore would produce much larger scatter.

The absence of significant relief effects is further indicated by the two traverses measured in two different inclusions. Although a change in the $\delta^{13}C_{PDB}$ values is observed from analytical spots close to the edge of the inclusions (Supplementary Figs 7 and 8), this change is not systematic, as might be expected from the relief-related fractionation effects. In the case of one inclusion, the $\delta^{13}C_{PDB}$ values near the edge are increasing, whereas in the second inclusion they are decreasing. In addition, the observed changes near the edges of inclusions are insufficient to change the average $\delta^{13}C_{PDB}$ values for these inclusions (Supplementary Figs 3 and 4). Therefore, it is apparent that the very light carbon compositions that are observed in most of the inclusions and are different from the bulk Earth value of about −6‰ cannot be produced by any instrument-related fractionation effects.

31. Farquhar, J., Hauri, E. & Wang, J. New insights into carbon fluid chemistry and graphite precipitation: SIMS analysis of granulite facies graphite from Ponmudi, South India. *Earth Planet. Sci. Lett.* **171,** 607–621 (1999).

LETTERS

Volcanic carbon dioxide vents show ecosystem effects of ocean acidification

Jason M. Hall-Spencer[1], Riccardo Rodolfo-Metalpa[1], Sophie Martin[2], Emma Ransome[1], Maoz Fine[3,4], Suzanne M. Turner[5], Sonia J. Rowley[1], Dario Tedesco[6,7] & Maria-Cristina Buia[8]

The atmospheric partial pressure of carbon dioxide (p_{CO_2}) will almost certainly be double that of pre-industrial levels by 2100 and will be considerably higher than at any time during the past few million years[1]. The oceans are a principal sink for anthropogenic CO_2 where it is estimated to have caused a 30% increase in the concentration of H^+ in ocean surface waters since the early 1900s and may lead to a drop in seawater pH of up to 0.5 units by 2100 (refs 2, 3). Our understanding of how increased ocean acidity may affect marine ecosystems is at present very limited as almost all studies have been *in vitro*, short-term, rapid perturbation experiments on isolated elements of the ecosystem[4,5]. Here we show the effects of acidification on benthic ecosystems at shallow coastal sites where volcanic CO_2 vents lower the pH of the water column. Along gradients of normal pH (8.1–8.2) to lowered pH (mean 7.8–7.9, minimum 7.4–7.5), typical rocky shore communities with abundant calcareous organisms shifted to communities lacking scleractinian corals with significant reductions in sea urchin and coralline algal abundance. To our knowledge, this is the first ecosystem-scale validation of predictions that these important groups of organisms are susceptible to elevated amounts of p_{CO_2}. Sea-grass production was highest in an area at mean pH 7.6 (1,827 μatm p_{CO_2}) where coralline algal biomass was significantly reduced and gastropod shells were dissolving due to periods of carbonate sub-saturation. The species populating the vent sites comprise a suite of organisms that are resilient to naturally high concentrations of p_{CO_2} and indicate that ocean acidification may benefit highly invasive non-native algal species. Our results provide the first *in situ* insights into how shallow water marine communities might change when susceptible organisms are removed owing to ocean acidification.

Short-term laboratory experiments show that many calcareous organisms may be unable to build their skeletons as oceans acidify over the next 100 years[6,7]. This may combine with other stresses, such as global warming, to drive tropical coral reefs towards functional collapse[8]. However, attempts to determine whether expectations on the basis of laboratory experiments and modelled predictions translate to field conditions have been hindered by the difficulty of imitating ocean acidification conditions *in situ* for sufficient periods to affect communities of macroorganisms.

Natural CO_2 flux from volcanic vents and high heat flow areas amounts to less than 0.5% of anthropogenic emissions to the global carbon budget, but can alter local ocean chemistry[9,10]. Marine CO_2 vents are abundant in the Mediterranean, especially around Italy and Greece where they typically eject volcanic fluids containing up to 1–2% hydrogen sulphide[10,11]. Some marine CO_2 vents are at ambient seawater temperature and lack toxic sulphur compounds; such vents can prevail for years to millenia[12] and may be used as natural experiments to advance our understanding of ocean acidification at the ecosystem level.

We studied cold vent areas off Ischia in Italy (Fig. 1) where sea water was being acidified by gas comprising 90.1–95.3% CO_2, 3.2–6.6% N_2, 0.6–0.8% O_2, 0.08–0.1% Ar and 0.2–0.8% CH_4 (no sulphur). Salinity (38%) and total alkalinity (2.5 mequiv. kg^{-1}) were homogeneous between survey stations and temperature-matched ambient seasonal fluctuations (13–25 °C). Vents occurred on the north and south sides of Castello d'Aragonese (40° 043.84′ N; 13° 57.08′ E) adjacent to a steeply sloping rocky shore. At the south vent site gas was emitted at 1.4×10^6 litre day^{-1} in an area of about 3,000 m^2 (mainly >5 vents m^{-2}); at the north site gas was emitted at 0.7×10^6 litre day^{-1} in an area of about 2,000 m^2 (mainly <5 vents m^{-2}). No seasonal, tidal or diurnal variation in gas flow rates was detected in 2006–07. The pH and saturation states (Ω) of calcite and aragonite varied with sea state, being lowest on calm days, and showed large decreases as p_{CO_2} amounts increased from approximately 300 to more than 2,000 μatm through the venting gas fields (Fig. 2 and Supplementary Table 2). Here we examine ecological tipping points along gradients of increasing p_{CO_2}, comparing normal pH stations (N_1, S_1 and P_1–P_2) with three stations that had reductions in mean pH of 0.2–0.4 units (N_2, S_2 and P_3; Fig. 1) and three stations (P_4, N_3 and S_3) with reductions in mean pH of 0.6–1.5 units which are more representative of the localized effects to be expected from deliberate CO_2 sequestration[13] rather than from global ocean acidification.

Rocky-shore stations with a mean pH of 7.8–7.9 (mean p_{CO_2} 804–957 μatm) showed a 30% reduction in species numbers (notably calcifiers) compared with the normal pH stations (Supplementary Tables 3 and 4). Temporal variability in p_{CO_2} will have contributed to the pronounced biodiversity shifts observed, as these stations experienced short periods of pH as low as 7.4–7.5. Organisms with aragonite skeletons were common outside the vents (for example, *Halimeda* algae and the corals *Caryophyllia*, *Cladocora* and *Balanophyllia*) but were absent at mean $\Omega_{arag} \leq 2.5$ (minimum Ω_{arag} 0.8–1.2), providing *in situ* support for predictions of global coral reef dissolution at these concentrations[8]. Although scleractinians can survive skeletal dissolution as polyps in the laboratory[14], reduced calcification due to low Ω_{arag} may result in increased risk to predation or competition in open ecosystems. The only Cnidaria in waters undersaturated with aragonite were anemones such as *Anemonia viridis*, which may benefit from increased p_{CO_2} for photosynthesis of its endosymbiotic dinoflagellates. Although atmospheric diffusion of CO_2 is not predicted to result in aragonite undersaturation in shallow waters of the Mediterranean,

[1]Marine Institute, Marine Biology and Ecology Research Centre, University of Plymouth, Plymouth PL4 8AA, UK. [2]CNRS-Université de Paris 6, Villefranche-sur-Mer 06234, France. [3]Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel. [4]The Interuniversity Institute for Marine Science, Eilat 88103, Israel. [5]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK. [6]Department of Environmental Sciences, 2nd University of Naples, Caserta 81100, Italy. [7]Istituto di Geologia Ambientale e Geoingegneria, CNR, Rome 00138, Italy. [8]Laboratorio di Ecologia del Benthos, Stazione Zoologica Anton Dohrn, Naples 80077, Italy.

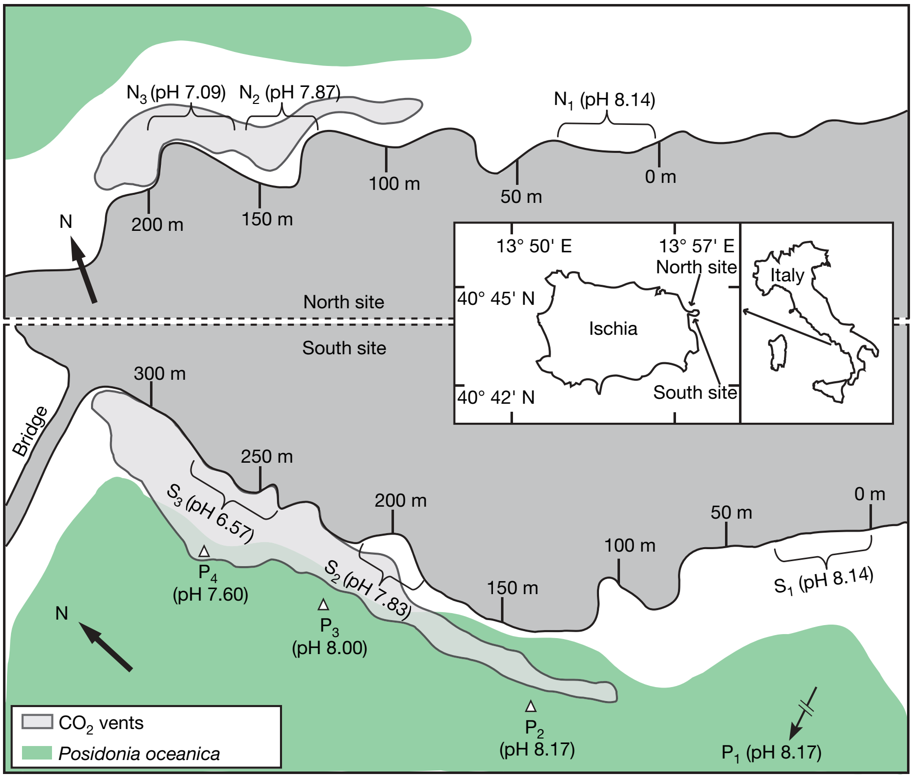


Figure 1 | Map of CO_2 vent sites north and south of Castello d'Aragonese, off Ischia Island, Italy. Mean surface pH is shown at 35-m-wide rocky-shore stations N_1–N_3 and S_1–S_3. Mean subtidal pH is shown at stations P_1–P_4, together with the distributions of CO_2 vents and *P. oceanica* sea-grass meadows. Reference station P_1 was at a 3-m depth, 400 m from the arrow shown.

observations of such areas are relevant to the localized effects caused by deliberate CO_2 sequestration and to the widespread effects predicted for areas that at present have low Ω_{arag}, given that high-latitude pteropods and coral reefs may be unable to make their skeletons by the year 2100 (refs 7, 13).

Mesocosm experiments have led to predictions that Corallinaceae, which help to protect against coral reef erosion in the tropics, are vulnerable to ocean acidification due to the solubility of their high magnesium calcite skeletons[15,16]. We found that Corallinaceae cover was significantly reduced at lowered pH (Table 1 and Supplementary Tables 2–4). As coralline algal cover fell from >60% outside the vent area to zero within it, non-calcareous algal cover increased significantly from near zero to >60% (Fig. 2 and Table 1). A suite of algal genera proved to be resilient to naturally high amounts of p_{CO_2} (for example, *Caulerpa*, *Cladophora*, *Asparagopsis*, *Dictyota* and *Sargassum*), some of which include invasive alien species that have begun to alter shallow marine ecosystems worldwide[17]. This adds to previously scant experimental information about the sorts of marine phototrophs that have enhanced growth and undiminished rates of photosynthesis at elevated concentrations of CO_2 (refs 4, 5, 18, 19).

The analysed *Posidonia oceanica* shoots were >10 yr old at the subtidal study sites and will have integrated the effects of lowered pH over this time. Sea-grass leaves at P_1 (pH 8.2) had 75% cover of calcified epiphytes but only 2% cover at P_4 (mean pH 7.6) with a significant reduction in epiphytic calcium carbonate per leaf (Table 1 and Figs 3 and 4). When heavily epiphytised leaves were transplanted from station P_1 to P_4 they showed complete dissolution of Corallinaceae in 2 weeks, whereas transplants moved within P_1 were unaffected. Mesocosm experiments have shown that sea-grass production can be enhanced at high p_{CO_2} (ref. 19). We found no difference (Table 1) in the photosynthetic performances of individual *P. oceanica* leaves between the four stations (mean ± s.e.m., photosynthetic efficiency (F_v/F_m) 0.74 ± 0.01 and electron transport rates $(ETR)_{max}$ 8.4 ± 1.9, $n = 40$) but sea-grass production was

Table 1 | Analysis of ecological tipping-points along marine acidity gradients

Category, site	*F* (d.f.)	*P* value	Tukey's test, site comparison
Corallinaceae cover, north	$F_{2,21} = 43.8$	0.000	$N_1 > N_2 > N_3$
Corallinaceae cover, south	$F_{2,21} = 48.0$	0.000	$S_1 > S_2 = S_3$
Non-calcareous crustose algal cover, north	$F_{2,21} = 0.31$	0.74	NS
Non-calcareous crustose algal cover, south	$F_{2,21} = 62.5$	0.000	$S_1 = S_2 < S_3$
Sea-grass epiphyte weight, south	$F_{3,315} = 176.2$	0.000	$P_1 > P_2 > P_3 > P_4$
Sea-grass F_v/F_m, south	$F_{3,36} = 0.13$	0.93	NS
Sea-grass ETR_{max}, south	$F_{3,36} = 0.06$	0.98	NS
Sea-grass shoot density, south	$F_{3,16} = 67.6$	0.000	$P_1 = P_2 = P_3 < P_4$
Sea urchin abundance, north	$F_{2,9} = 14.7$	0.001	$N_1 > N_2 = N_3$
Sea urchin abundance, south	$F_{2,9} = 65.3$	0.000	$S_1 > S_2 = S_3$
C. stellatus abundance, north	$F_{2,21} = 0.72$	0.50	NS
C. stellatus abundance, south	$F_{2,21} = 29.4$	0.000	$S_1 = S_2 > S_3$
O. turbinata abundance, north	$F_{2,21} = 3.50$	0.049	$N_1 = N_2 > N_3$
O. turbinata abundance, south	$F_{2,21} = 6.39$	0.007	$S_1 = S_3 < S_2$
P. caerulea abundance, north	$F_{2,21} = 22.8$	0.000	$N_1 > N_2 > N_3$
P. caerulea abundance, south	$F_{2,21} = 9.24$	0.001	$S_1 = S_2 > S_3$

Significant differences were assessed using one-way analysis of variance (ANOVA, *F*) and Tukey's HSD (honestly significant difference) post-hoc tests. Data are from stations north and south of Castello d'Aragonese, Ischia, Italy in spring 2007. d.f., degrees of freedom, NS, not significant.

highest at mean pH 7.6 (biomass increased by 2.8 g m^{-2} day^{-1} at mean p_{CO_2} 1,827 μatm) where shoot density was significantly higher (Table 1 and Fig. 3) and approximately 30% higher than that known anywhere else around Ischia[12].

Sea urchins (*Paracentrotus lividus*, *Arbacia lixula*), which have high magnesium calcite skeletons, were the most common large invertebrates on sublittoral rock outside the vents but their abundance was significantly reduced where pH reached minima of 7.4–7.5 (Table 1 and Fig. 2). This supports physiological studies showing that sea urchins are vulnerable to a rise in CO_2, and is a concern as sea urchin loss can drive deteriorations in ecosystem complexity and stability[20,21]. Although sea urchins cannot close off their supply of ambient sea water, some organisms can do this to avoid pH minima. Other calcitic organisms, such as the barnacle *Chthamalus stellatus*, for example, may survive pH minima by closing their rostral plates as

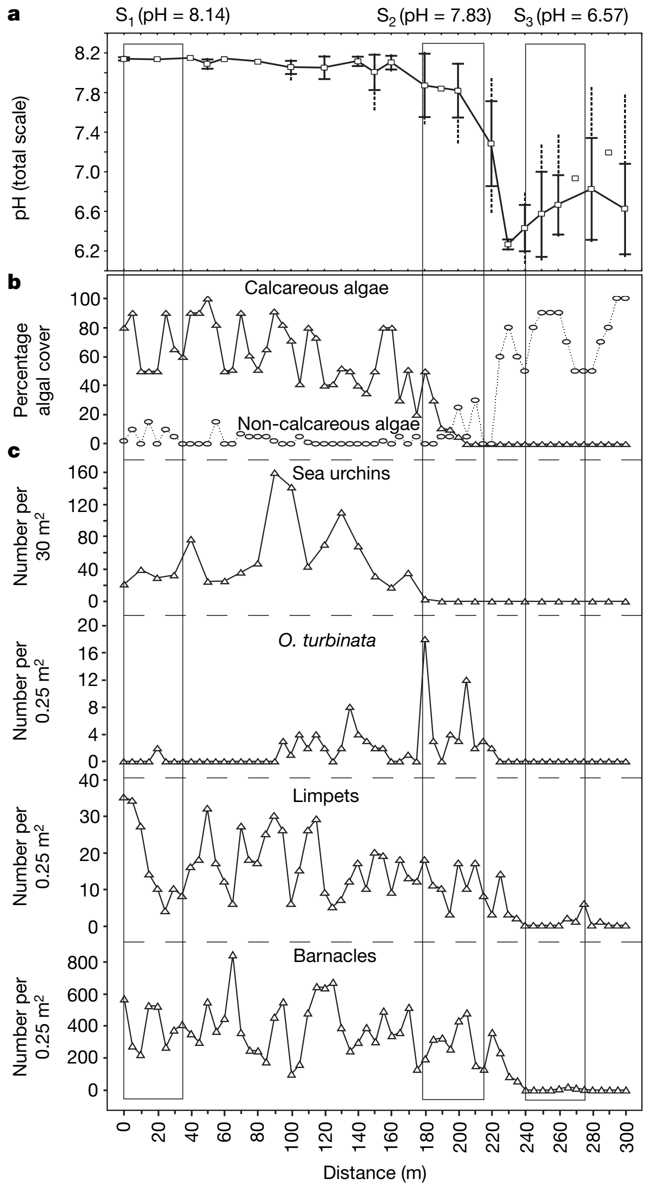


Figure 2 | Variation in pH, cover of algae and abundance of species at CO_2 vents south of Castello d'Aragonese. Data are from stations S_1–S_3 (see Fig. 1) from 18 April to 9 May 2007. **a**, The mean pH ± s.d. (cross bars) is shown. Ranges are denoted by the dotted line; $n = 6$ at 0 m, $n = 11$ at 50 m, 100 m, 250 m and 300 m, $n = 9$ at 220 m, 260 m, 280 m and $n = 12$ at 150 m and 200 m. **b**, The percentage cover of calcareous (triangles) and non-calcareous algae (circles) is shown. **c**, The abundances of sea urchins, *O. turbinata*, limpets and barnacles.

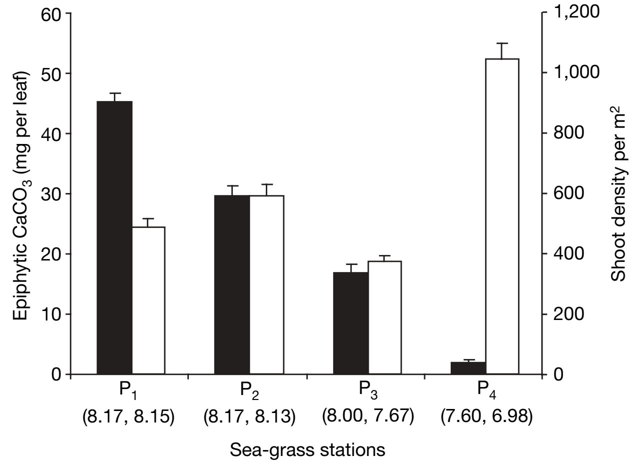


Figure 3 | Sea-grass shoot density and amount of epiphytic $CaCO_3$ on leaves growing at differing pH levels south of Castello d'Aragonese. Shoot density (open column, $n = 4$, mean and s.d.) and epiphytic $CaCO_3$ (filled column, $n = 80$, mean and s.d.) for data from 18 April to 9 May 2007 at various pH levels (mean and minimum values are shown; P_1 $n = 30$, P_2 $n = 16$, P_3 $n = 23$ and P_4 $n = 37$).

their abundance was not significantly reduced until extremely low mean pH 6.6 (Table 1 and Fig. 2). Juveniles of *Osilinus turbinata* and *Patella caerulea* gastropods were absent in areas with pH minima ≤7.4, where all adult gastropod shells (including *Hexaplex trunculus* and *Cerithium vulgatum*) were weakened by the acidified sea water (Figs 2 and 4, Table 1 and Supplementary Video), an effect which probably increases their risk of predation[22].

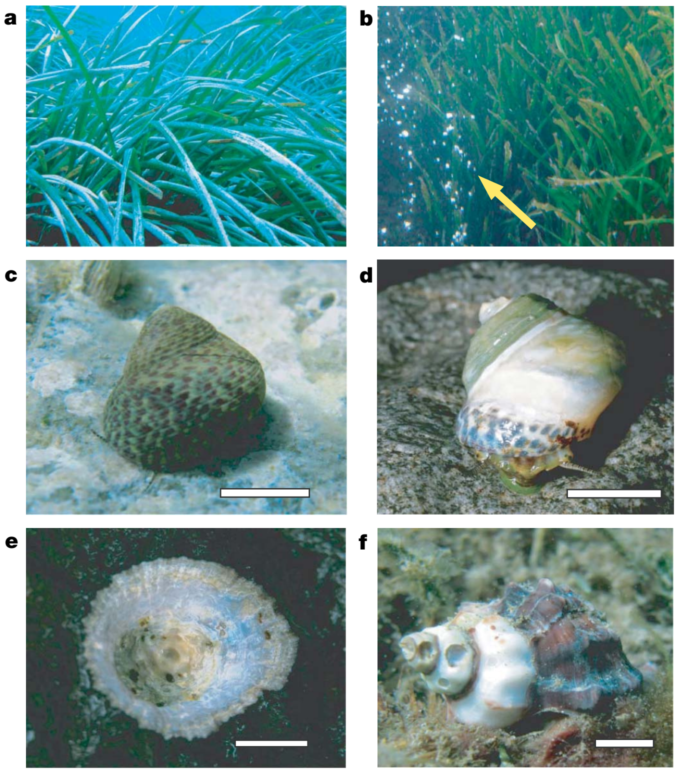


Figure 4 | Dissolution of calcified organisms due to naturally acidified sea water. **a**, **b**, *Posidonia oceanica* with heavy overgrowth of Corallinaceae at pH 8.2 (**a**) and lacking Corallinaceae at mean pH 7.6 (**b**); arrow indicates bubbles from the CO_2 vent field. **c**, **d**, Typical examples of *O. turbinata* with the periostracum intact at pH 8.2 (**c**) and with old parts of the periostracum removed at mean pH 7.3 (**d**). **e**, **f**, Live *P. caerulea* (**e**) and *H. trunculus* (**f**) showing severely eroded, pitted shells in areas of minimum pH 7.4. Scale bars represent 1 cm.

Vent systems are not perfect predictors of future ocean ecology owing to temporal variability in pH, spatial proximity of populations unaffected by acidification and the unknown effects of other global changes in parameters such as temperature, currents and sea level. However, such vents acidify sea water on sufficiently large spatial and temporal scales to integrate ecosystem processes such as production, competition and predation. Lush stands of sea-grass and brown algae can thrive along natural pH gradients where aragonitic and then calcitic calcareous organisms are lost owing to skeletal dissolution. This confirms experimental and modelling predictions that differential responses of benthic species to decreased pH can lead to substantial changes in community structure[4–8,13–16]. Many of the organisms that were adversely affected by reductions in pH at our study sites belong to groups that existed before and after periods of similar reductions in the past (for example, calcified algae, corals and sea urchins)[14]. It is unknown whether there will be sufficient refugia or enough time for these groups to adapt to survive the rapid rate of ocean acidification predicted due to anthropogenic CO_2. This opportunity to observe the tipping points at which principal groups of marine organisms are affected by lowered pH proves that, even without global warming, the projected rise in atmospheric CO_2 concentration is hazardous, as ocean acidification will probably bring about reductions in biodiversity and radically alter ecosystems.

METHODS SUMMARY

Vent gases were collected in pre-evacuated glass flasks partly filled with 0.1 M $Cd(OH)_2$ and 4 N NaOH solution (see Supplementary Video). Uncondensable gases were collected in the headspace, inorganic residual gas compounds were analysed using thermal conductivity chromatographs, methane was analysed with a flame ionization detector and ion chromatography was used to analyse condensable gases such as CO_2 dissolved during collection. Between 18 April and 9 May 2007, surface and bottom water samples were regularly taken for measurements of the spatial and temporal variability in pH (in total scale), total alkalinity and salinity in various weather conditions. In winter 2006, and spring and autumn 2007, intertidal and subtidal SCUBA surveys were made of the main macroorganisms present within and adjacent to the vents to 3 m depth. Epibiont calcium carbonate on *P. oceanica* leaves was quantified along a gradient of pH; leaves that were heavily encrusted with Corallinaceae were transplanted from a reference site into an area with mean pH 7.6 then reassessed after 2 weeks. *Posidonia oceanica* production, growth dynamics and shoot density was estimated at stations P_1–P_4 where their photosynthetic efficiency (F_v/F_m) and electron transport rates (ETR) were measured *in situ* using a diving pulse amplitude modulation (PAM), and in the laboratory using an imaging PAM.

Received 13 March; accepted 1 May 2008.
Published online 8 June 2008.

1. Pearson, P. N. & Palmer, M. R. Atmospheric carbon dioxide concentrations over the past 60 million years. *Nature* **406**, 695–699 (2000).
2. Intergovernmental Panel on Climate Change. Summary for Policymakers. In *Climate Change 2007: The Physical Sciences Basis. Working Group I Contribution to the Fourth Assessment Report of the IPCC* (eds Solomon, S. *et al.*) (Cambridge Univ. Press, Cambridge, 2007).
3. Caldeira, K. & Wickett, M. E. Ocean model predictions of chemistry changes from carbon dioxide emissions to the atmosphere and ocean. *J. Geophys. Res.* **110**, C09S04, doi:10.1029/2004JC002671 (2005).
4. The Royal Society. Ocean acidification due to increasing atmospheric carbon dioxide. Policy document 12/05 (The Royal Society, London, 2005).
5. Riebesell, U. *et al.* Enhanced biological carbon consumption in a high CO_2 ocean. *Nature* **450**, 545–548 (2007).
6. Feely, R. A. *et al.* Impact of anthropogenic CO_2 on the $CaCO_3$ system in the oceans. *Science* **305**, 362–366 (2004).
7. Orr, J. C. *et al.* Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms. *Nature* **437**, 681–686 (2005).
8. Hoegh-Guldberg, O. *et al.* Coral reefs under rapid climate change and ocean acidification. *Science* **318**, 1737–1742 (2007).
9. Kerrick, D. M., McKibben, M. A., Seward, T. M. & Caldeira, K. Convective hydrothermal CO_2 emission from high heat-flow regions. *Chem. Geol.* **121**, 285–293 (1995).
10. Williams, S. N., Schaefer, S. J., Calvache v., M. L. & Lopez, D. Global carbon dioxide emission to the atmosphere by volcanoes. *Geochim. Cosmochim. Acta* **56**, 1765–1770 (1992).
11. Dando, P. R., Stuben, D. & Varnavas, S. P. Hydrothermalism in the Mediterranean Sea. *Prog. Oceanogr.* **44**, 333–367 (1999).
12. Ambiente. *Marino Costiero e Territorio Delle Isole Flegree* (eds Gambi, M. C., Lauro, M. & Jannuzzi, F.) (Accademia di Scienze Fische e Matematiche, Italy, 2003).
13. Davies, A. J., Roberts, J. M. & Hall-Spencer, J. Preserving deep-sea natural heritage: emerging issues in offshore conservation and management. *Biol. Conserv.* **138**, 299–312 (2007).
14. Fine, M. & Tchernov, D. Scleractinian coral species survive and recover from decalcification. *Science* **315**, 1811 (2007).
15. Kleypas, J. A. *et al. Impacts of Ocean Acidification on Coral Reefs and Other Marine Calcifiers: A Guide for Future Research.* Report of a workshop held 18–20 April 2005, St. Petersburg, FL, sponsored by NSF, NOAA, and the US Geological Survey (2006).
16. Kuffner, I. B. *et al.* Decreased abundance of crustose coralline algae due to ocean acidification. *Nature Geosci* **1**, 114–117 (2008).
17. Boudouresque, C. F. & Verlaque, M. Biological pollution in the Meditterranean Sea: invasive versus introduced macrophytes. *Mar. Pollut. Bull.* **44**, 32–38 (2002).
18. Levitan, O. *et al.* Elevated CO_2 enhances nitrogen fixation and growth in the marine cyanobacterium *Trichodesmium. Glob. Change Biol.* **13**, 531–538 (2007).
19. Palacios, S. L. & Zimmerman, R. C. Response of eelgrass *Zostera marina* to CO_2 enrichment: possible impacts of climate change and potential for remediation of coastal habitats. *Mar. Ecol. Prog. Ser.* **344**, 1–13 (2007).
20. Miles, H., Widdicombe, S., Spicer, J. I. & Hall-Spencer, J. M. Effects of anthropogenic seawater acidification on acid-base balance in the sea urchin *Psammechinus miliaris. Mar. Pollut. Bull.* **54**, 89–96 (2007).
21. Scheffer, M. *et al.* Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).
22. Bibby, R. *et al.* Ocean acidification disrupts induced defences in the intertidal gastropod *Littorina littorea. Biol. Lett.* **3**, 699–701 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the staff of Anton Dohrn Benthic laboratory, Ischia for technical help. J.M.H.-S. was funded by a Royal Society University Research Fellowship and was first shown the gas vent sites by M. Taviani in 2002; R.R.-M. and S.M.T. were funded by the Leverhulme Trust. A. de Simone, A. Ferrara and M. Laurenti helped with field measurements, V. King took photo 4d, and O. Hoegh Guldberg and P. Liss helped improve the manuscript.

Author Contributions All authors were involved with fieldwork and sample analyses. J.M.H.-S. designed the study and wrote the paper along with R.R.-M., M.F. and S.M.T. D.T. analysed gases, S.M. analysed sea-grass epiphytes and seawater chemistry, E.R. and S.J.R. collected intertidal and subtidal data respectively, and M.-C.B. provided sea-grass expertise. All authors discussed results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.H.-S. (jhall-spencer@plymouth.ac.uk).

LETTERS

Extinction risk depends strongly on factors contributing to stochasticity

Brett A. Melbourne[1] & Alan Hastings[2]

Extinction risk in natural populations depends on stochastic factors that affect individuals, and is estimated by incorporating such factors into stochastic models[1–9]. Stochasticity can be divided into four categories, which include the probabilistic nature of birth and death at the level of individuals (demographic stochasticity[2]), variation in population-level birth and death rates among times or locations (environmental stochasticity[1,3]), the sex of individuals[6,8] and variation in vital rates among individuals within a population (demographic heterogeneity[7,9]). Mechanistic stochastic models that include all of these factors have not previously been developed to examine their combined effects on extinction risk. Here we derive a family of stochastic Ricker models using different combinations of all these stochastic factors, and show that extinction risk depends strongly on the combination of factors that contribute to stochasticity. Furthermore, we show that only with the full stochastic model can the relative importance of environmental and demographic variability, and therefore extinction risk, be correctly determined. Using the full model, we find that demographic sources of stochasticity are the prominent cause of variability in a laboratory population of *Tribolium castaneum* (red flour beetle), whereas using only the standard simpler models would lead to the erroneous conclusion that environmental variability dominates. Our results demonstrate that current estimates of extinction risk for natural populations could be greatly underestimated because variability has been mistakenly attributed to the environment rather than the demographic factors described here that entail much higher extinction risk for the same variability level.

An essential question in ecology and conservation biology is the determination of the likelihood of extinction within a biological system[10]. This clearly depends on understanding the relative importance of different processes that affect the stochastic dynamics of biological populations, and how these interact with both density-dependent and density-independent processes[5,6]. Ecologists have long sought simple approaches to predicting the likelihood of extinction[11,12]. In conservation biology, the simple idea of a population level that determines which kind of forces might lead to extinction has been appealing[4,13–15]. However, a more detailed and more mechanistic approach is clearly needed to answer these questions more carefully in a way that uses available data.

There is a long history of models that incorporate stochasticity to examine its effect on population growth and extinction[1–6,13,16–21]. The first stochastic models showed that populations could become extinct even if deterministic models concluded they would persist indefinitely[16]. Early results also showed that the variance of population fluctuations and the probability of extinction depend on which biological processes are subject to stochasticity, and that the long-term growth rate of a stochastic population differs from an equivalent population with deterministic dynamics[16,17]. These general results have proved to be robust, and later studies have concentrated on how different sources of stochasticity in the life history of organisms affect population growth and extinction.

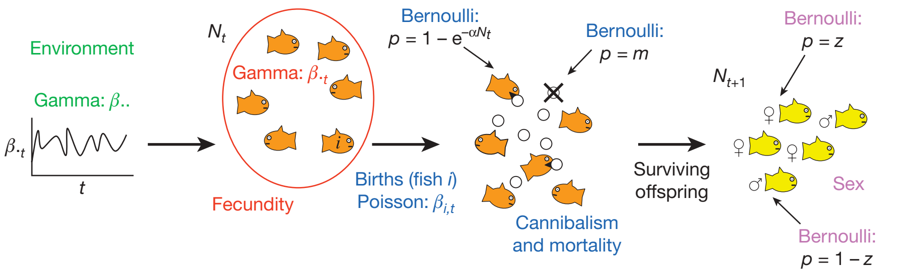
There are many sources of stochasticity that contribute to variance in population growth and thus contribute to the risk of stochastic extinction. Two broad classes are most commonly recognized[6]. Demographic stochasticity occurs because the birth or death of an individual is a random event, such that individuals identical in their probability distributions for reproduction or longevity nevertheless differ by chance in how many offspring they produce or when they will die[2,20]. Environmental stochasticity occurs because fluctuations in exogenous environmental factors such as temperature and rainfall drive population-level fluctuations in birth and death rates[3,20]. In small populations, demographic stochasticity increases extinction risk due to unfortunate coincidences in the fate of individuals, which are cancelled out in larger populations. In contrast, environmental stochasticity increases extinction risk over a larger range of population sizes because the whole population is affected simultaneously.

Two further sources of stochasticity have long been recognized[17] but only recently analysed, namely stochastic sex determination[6,8,22,23] and demographic heterogeneity[7,9], with the former strictly an extreme form of the latter. These can both be viewed as components of demographic stochasticity[6,7], although we separate them here because they are fundamentally different to randomness in births and deaths. In sexually reproducing species, the sex of an offspring is often randomly determined, giving rise to a stochastically fluctuating sex ratio in the population. Most current models of extinction risk only include females; however, a stochastic sex ratio can increase the variance in population growth and extinction risk over and above the effects of demographic stochasticity on females alone. This is because males contribute to density-dependent regulation or because the lack of males reduces female mating success[8,23,24].

Demographic heterogeneity refers to variation in birth or death rates among individuals within a population, such as might occur among individuals of different size[7,9]. This contrasts with demographic stochasticity, which in its original definition and subsequent application concerns chance events assuming a fixed value of the birth or death rate of an individual[2,20]. Demographic stochasticity, sex ratio stochasticity and demographic heterogeneity all contribute to the total demographic variance. Demographic heterogeneity can either increase or decrease the demographic variance, depending on the details of the stochastic process, and so can either increase or decrease the extinction risk[7].

A problem that remains to be addressed is how to combine the various sources of stochasticity into an analytically tractable model. Many current approaches begin by assuming a deterministic skeleton to which noise terms are added, where the statistical distribution of the noise is chosen to reflect a broad class of stochasticity[6,25]. Among

[1]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA. [2]Department of Environmental Science and Policy, University of California, Davis, California 95616, USA.



Demographic stochasticity	Environmental stochasticity	Demographic heterogeneity	Sex	Model	Abbreviation
●				Poisson	P
●	●			Negative binomial environmental	NBe
●		●		Negative binomial demographic	NBd
●	●	●		Negative binomial gamma	NBg
●			●	Poisson binomial	PB
●	●		●	Negative binomial-binomial environmental	NBBe
●		●	●	Negative binomial-binomial demographic	NBBd
●	●	●	●	Negative binomial-binomial gamma	NBBg

Figure 1 | A family of stochastic Ricker models based on Ricker's[26] assumptions about the life cycle of a fish species that cannibalises its eggs. The stochastic models incorporate stochasticity in various parts of the life cycle, including gamma variation in environmentally determined birth rates, gamma variation in birth rates between individuals, Poisson variation in birth rates within individuals, Bernoulli variation in mortality within individuals, and Bernoulli variation in the sex of an individual at birth.

other models, the Ricker model[26] has often been used as a deterministic skeleton[25,27]. Here we incorporate stochasticity directly into the birth and death processes, allowing the mean and variance of population growth to arise mechanistically from the underlying process assumptions. Our models are for discrete individuals. We derive our stochastic models from Ricker's assumptions but extend these by specifying the stochastic mechanisms at different stages in the life history of an individual and scaling up to the population level (Supplementary Methods). Ricker's assumptions[26] lead to the Poisson–Ricker model, which contains demographic stochasticity arising from the number of eggs laid by individuals and the survival of individual eggs from predation by adults. To this basic model we add environmental stochasticity and demographic heterogeneity in the number of offspring and stochasticity in the sex of offspring. We focus on births because variability in births has greater or equal effects than mortality, but our models extend generally to mortality variation (Supplementary Discussion). We use different combinations of the various stochastic sources to derive a family of nested stochastic Ricker models (Fig. 1).

The stochastic models are true Ricker models because they all have conditional mean N_{t+1} equal to the deterministic Ricker model[26], that is, $E[N_{t+1}] = RN_t\exp(-\alpha N_t)$, where N_t is the population size in generation t, R is the density-independent mean per capita growth rate (finite rate of growth), and α is a measure of density-dependent effects (Supplementary Methods). However, the various stochastic models have different distributions of numbers next year as a function of numbers this year (Supplementary Table 1) and so differ substantially in their variance characteristics for the number of individuals in a subsequent generation (Fig. 2 and Supplementary Fig. 1). As expected, the variance in the number of individuals in the next generation increases as more sources of stochasticity are included in the models. The Poisson–Ricker model, a model of pure demographic stochasticity, has the smallest variance (Fig. 2).

When the total variance is held at the same value (Supplementary Methods), there is an important difference between models of environmental stochasticity and demographic heterogeneity in the variance for the number of individuals the following generation (Fig. 2). For environmental stochasticity, the variance in numbers peaks at the stationary point of the deterministic Ricker function, whereas for demographic heterogeneity, the variance is concentrated at low abundance to the left of the stationary point. This is because environmental stochasticity results in a density-independent variance parameter, whereas demographic heterogeneity generates one that is density dependent (Supplementary Methods). Consequently, demographic heterogeneity entails a greater risk of extinction than environmental stochasticity for the same total variance (Fig. 3). As we highlight below, the similarities in the two variance functions allow these processes to be easily confused, yet their differences have large effects on extinction risk.

The stochastic sex ratio increases the variance at low to intermediate initial abundance, and substantially so at abundances less than the stationary point of the Ricker model (Fig. 2). The effect of the sex ratio is greatest in the demographic models (Fig. 2; compare Poisson (P) with Poisson binomial (PB) models, and negative binomial

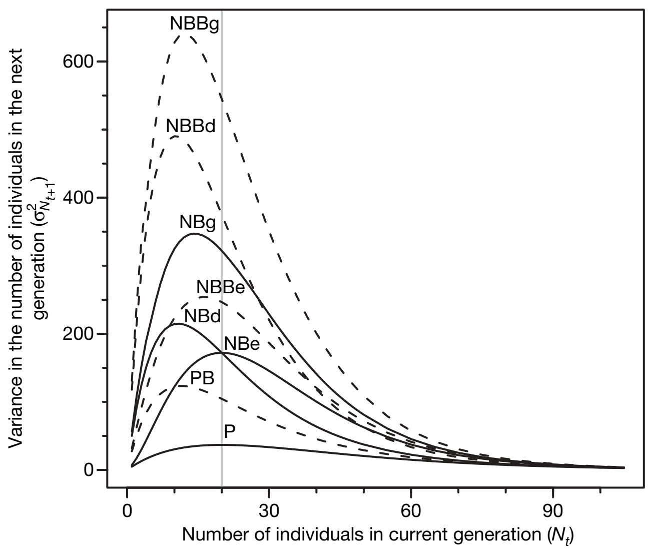


Figure 2 | Variance in the number of individuals in the next generation (N_{t+1}) as a function of the number of individuals in the current generation (N_t) for the stochastic Ricker models. The model parameters were: $R = 5$, $\alpha = 0.05$, $k_D = 0.5$, $k_E = 10$. The stochastic parameters (k_D, k_E) were set so that the total variance due to demographic heterogeneity was equal to the total variance due to environmental stochasticity. The vertical bar indicates the position of the stationary point in the Ricker production function. Abbreviations identify the models listed in Fig. 1.

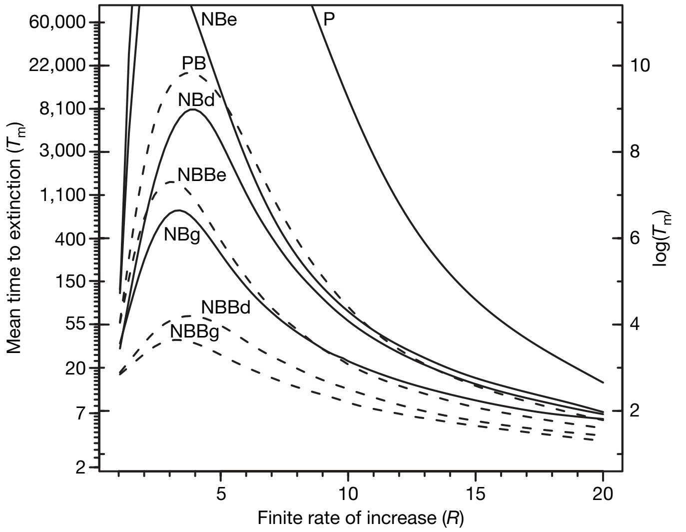


Figure 3 | Intrinsic mean time to extinction[30] (T_m) for the stochastic Ricker models as a function of the finite rate of increase (*R*). The model parameters were: $k_D = 0.5$; k_E was adjusted so that the total variance in N^{t+1} due to demographic heterogeneity was equal to the total variance due to environmental stochasticity; α was adjusted to hold the equilibrium density at 30 individuals. Internal ticks show the natural logarithm scale. Abbreviations identify the models listed in Fig. 1.

demographic (NBd) with negative binomial-binomial demographic (NBBd) models). The combined variance of demographic stochasticity, environmental stochasticity, demographic heterogeneity and stochastic sex ratio is higher than in models of their individual effects and is additive (Fig. 2).

Extinction risk for the stochastic Ricker models differs substantially depending on the combination of factors in the life cycle that contribute to stochasticity (Fig. 3). The lowest extinction risk is for the Poisson–Ricker model, which includes only demographic stochasticity, whereas the highest extinction risk is for the model that includes all sources of stochasticity. Notably, for the same total variance, extinction risk is enhanced more by demographic heterogeneity or a stochastic sex ratio than by environmental stochasticity, often by orders of magnitude. Extinction risk is also dependent on the finite rate of growth, *R* (Fig. 3). Increasing *R* from 1 initially promotes higher persistence times but it also enhances the contribution of nonlinear dynamics to the variance in population fluctuations, causing persistence times to eventually fall. For populations with growth rates *R* larger than the value producing the first bifurcation in the Ricker model (7.4), fluctuations due to nonlinear dynamics increase and persistence times rapidly drop below those of populations with *R* equal to 1 (the minimum *R* required for persistence in the absence of fluctuations).

The characteristic probability mass functions (Supplementary Table 1) of the different stochastic Ricker models provide an opportunity to distinguish between models by fitting them to data. Using likelihood approaches and information criteria[28], we fitted the models to data from a laboratory experiment on *Tribolium castaneum* growing in discrete time cultures in temperature-controlled incubators. As in Ricker's models of a fish species (Fig. 1), cannibalism of eggs by adults is the main density-regulating process in laboratory populations of *T. castaneum* in discrete time cultures[29]. The best-fitting model was the negative binomial-binomial gamma model, which is the only model to include all four sources of stochasticity (Table 1; the fitted model is shown in Supplementary Fig. 2). No other model fitted as well (Table 1) and the experimental design provided a robust distinction between the models (Supplementary Discussion). In addition, the second-best model (also by a substantial amount) was the negative binomial gamma model, which left out only the stochastic sex ratio that is then partly absorbed by the demographic heterogeneity parameter (Table 1).

The likelihood analysis revealed several important features of the stochastic system. The Poisson model was the worst model by a large margin (Table 1, ΔAIC = 336), suggesting that the most basic assumptions of demographic stochasticity in births, density-dependent survival and density-independent survival are completely unable to describe the variance in abundance even when environmental variability is tightly controlled within the laboratory. In addition, the estimated vital rates of the population were not very different among the models but the estimates of the stochastic parameters were very sensitive to which stochastic factors were included in the fitted model (Table 1). This highlights the importance of a full model specification for correctly identifying the important stochastic factors and therefore correctly estimating extinction risk. Notably, the full model revealed that demographic heterogeneity was much more important than environmental stochasticity, whereas simpler models without demographic heterogeneity erroneously suggest that environmental variability dominates because any demographic heterogeneity is absorbed by the environmental variance parameter (Table 1).

These results show that many species currently viewed as at risk of extinction from environmental stochasticity could instead be at much higher risk from undetected demographic variance. This demographic variance is driven by sex ratio variation and demographic heterogeneity that has been mistakenly attributed to environmental stochasticity. The increased extinction risk is a consequence of the fact that, for the same overall level of variance in abundance for one generational step, sex ratio stochasticity and demographic heterogeneity give rise to greater variance than environmental stochasticity when population sizes are small and vulnerable. Thus, identifying the relative contribution of different stochastic processes is vital to understanding fluctuations and estimating extinction risk because variability differs at different population levels for different processes. As natural populations are likely to have greater demographic heterogeneity than our laboratory stock of *T. castaneum*, the effect we have uncovered here will be larger in natural populations. Suitable data could include time series of population abundance using the methods we have developed here, or individual level data, with special effort needed to encompass a range of population

Table 1 | Fit of stochastic Ricker models to *T. castaneum* data

Model	R	α	k_D	k_E	L	ΔAIC
Poisson	2.526	0.003636	-	-	−406.5	336
Negative binomial demographic	2.638	0.003744	0.1463	-	−246.3	18
Negative binomial environmental	2.706	0.003800	-	1.9913	−265.3	56
Negative binomial gamma	2.598	0.003727	0.2610	29.2262	−238.9	5
Poisson binomial	2.697	0.003753	-	-	−282.0	87
Negative binomial-binomial demographic	2.621	0.003731	0.3876	-	−245.8	17
Negative binomial-binomial environmental	2.770	0.003831	-	13.1014	−242.6	10
Negative binomial-binomial gamma	2.613	0.003731	1.1475*	26.6221*	−236.4	0

The models were fitted to the data by maximizing the log likelihood (L), calculated from the probability mass function of each stochastic Ricker model (Supplementary Table 1). The estimated parameters were: R, the density-independent mean per capita growth rate; α, the density-dependent parameter; k_D and k_E, the variance parameters for demographic heterogeneity and environmental stochasticity, respectively, where small values indicate large variance. The difference in the Akaike information criterion (ΔAIC) was used to compare models[28].
* Bias-corrected estimates for k_D and k_E were 1.07 and 17.62, respectively (see Supplementary Discussion).

densities to capture the density-dependent nature of the variance in abundance. With field data, care will also be needed to factor in measurement error as this will further hide the importance of demographic heterogeneity relative to environmental stochasticity (Supplementary Discussion). We suggest that extinction risk for many populations of conservation concern needs to be urgently re-evaluated with full consideration of all factors contributing to stochasticity.

METHODS SUMMARY

We placed adult *T. castaneum* into 4 cm × 4 cm × 6 cm acrylic containers with 20 g of standard medium (95% flour, 5% brewer's yeast) to lay eggs for 24 h, after which time the adults were removed. We set up 60 separate containers with adult numbers ranging from 2 to 1,000. Containers were kept in a constant-temperature incubator at 31 °C for the full beetle life cycle and their positions within the incubator were randomized weekly. The 24-h egg-laying period was followed by a further 34 days during which individuals passed through the egg, larval and pupal stages. The number of adults emerging at the end of the 35-day life cycle was recorded for each container. The stochastic Ricker models were fitted to the emergence data by maximum likelihood[28].

Received 10 December 2007; accepted 13 March 2008.

1. Athreya, K. B. & Karlin, S. On branching processes with random environments: extinction probabilities. *Ann. Math. Stat.* **42,** 1499–1520 (1971).
2. May, R. M. *Stability and Complexity in Model Ecosystems* (Princeton Univ. Press, Princeton, 1973).
3. May, R. M. Stability in randomly fluctuating versus deterministic environments. *Am. Nat.* **107,** 621–650 (1973).
4. Lande, R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* **142,** 911–927 (1993).
5. Ludwig, D. The distribution of population survival times. *Am. Nat.* **147,** 506–526 (1996).
6. Lande, R., Engen, S. & Saether, B. E. *Stochastic Population Dynamics in Ecology and Conservation* (Oxford Univ. Press, Oxford, UK, 2003).
7. Kendall, B. E. & Fox, G. A. Unstructured individual variation and demographic stochasticity. *Conserv. Biol.* **17,** 1170–1172 (2003).
8. Saether, B. E. *et al.* Time to extinction in relation to mating system and type of density regulation in populations with two sexes. *J. Anim. Ecol.* **73,** 925–934 (2004).
9. Fox, G. A., Kendall, B. E., Fitzpatrick, J. W. & Woolfenden, G. E. Consequences of heterogeneity in survival probability in a population of Florida scrub-jays. *J. Anim. Ecol.* **75,** 921–927 (2006).
10. Soulé, M. E. (ed.) *Viable Populations for Conservation* (Cambridge Univ. Press, Cambridge, UK, 1987).
11. Shaffer, M. L. Minimum population sizes for species conservation. *Bioscience* **31,** 131–134 (1981).
12. Pimm, S. L., Jones, H. L. & Diamond, J. On the risk of extinction. *Am. Nat.* **132,** 757–785 (1988).
13. Leigh, E. G. The average lifetime of a population in a varying environment. *J. Theor. Biol.* **90,** 213–239 (1981).
14. Goodman, D. in *Viable Populations for Conservation* (ed. Soulé, M. E.) 11–34 (Cambridge Univ. Press, Cambridge, UK, 1987).
15. Morris, W. F. & Doak, D. F. *Quantitative Conservation Biology: Theory and Practice of Population Viability Analysis* (Sinauer Associates Inc., Sunderland, Massachusetts, 2002).
16. Feller, W. Die Grundlagen der Volterraschen Theorie des Kampfes ums Dasein in wahrscheinlichkeitstheoretischer Behandlung. *Acta Biotheor.* **5,** 11–40 (1939).
17. Kendall, D. G. Stochastic processes and population growth. *J. R. Stat. Soc. Ser. B Methodol.* **11,** 230–282 (1949).
18. Bartlett, M. S. *Stochastic Population Models in Ecology and Epidemiology* (Methuen, London, 1960).
19. Lewontin, R. C. & Cohen, D. On population growth in a randomly varying environment. *Proc. Natl Acad. Sci. USA* **62,** 1056–1060 (1969).
20. Roughgarden, J. A simple model for population dynamics in stochastic environments. *Am. Nat.* **109,** 713–736 (1975).
21. Tuljapurkar, S. An uncertain life: demography in random environments. *Theor. Popul. Biol.* **35,** 227–294 (1989).
22. Gabriel, W. & Burger, R. Survival of small populations under demographic stochasticity. *Theor. Popul. Biol.* **41,** 44–71 (1992).
23. Engen, S., Lande, R. & Saether, B. E. Demographic stochasticity and Allee effects in populations with two sexes. *Ecology* **84,** 2378–2386 (2003).
24. Legendre, S., Clobert, J., Moller, A. P. & Sorci, G. Demographic stochasticity and social mating system in the process of extinction of small populations: The case of passerines introduced to New Zealand. *Am. Nat.* **153,** 449–463 (1999).
25. Dennis, B., Desharnais, R. A., Cushing, J. M., Henson, S. M. & Costantino, R. F. Estimating chaos and complex dynamics in an insect population. *Ecol. Monogr.* **71,** 277–303 (2001).
26. Ricker, W. E. Stock and recruitment. *J. Fish. Res. Bd Can.* **11,** 559–623 (1954).
27. Drake, J. M. Density-dependent demographic variation determines extinction rate of experimental populations. *PLoS Biol.* **3,** e222 (2005).
28. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* (Springer, New York, 2002).
29. Costantino, R. F. & Desharnais, R. A. *Population Dynamics and the* Tribolium *Model: Genetics and Demography* (Springer, New York, 1991).
30. Grimm, V. & Wissel, C. The intrinsic mean time to extinction: a unifying approach to analysing persistence and viability of populations. *Oikos* **105,** 501–511 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Gibson, D. Hodgkiss, C. Koenig, T. McCabe, D. Paulus, D. Smith, N. Tcheou, R. Villalobos and M. Wu for assistance. This study was funded by the National Science Foundation.

Author Contributions B.A.M. derived and analysed the models, and analysed the data. B.A.M. and A.H. conceived the study, planned and directed the experiments, and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to B.A.M. (brett.melbourne@colorado.edu).

LETTERS

A myocardial lineage derives from *Tbx18* epicardial cells

Chen-Leng Cai[1,4,5,6]*, Jody C. Martin[2]*, Yunfu Sun[1]*, Li Cui[1], Lianchun Wang[7], Kunfu Ouyang[3], Lei Yang[1], Lei Bu[1], Xingqun Liang[3], Xiaoxue Zhang[1], William B. Stallcup[8], Christopher P. Denton[9], Andrew McCulloch[2], Ju Chen[3] & Sylvia M. Evans[1,3]

Understanding the origins and roles of cardiac progenitor cells is important for elucidating the pathogenesis of congenital and acquired heart diseases[1,2]. Moreover, manipulation of cardiac myocyte progenitors has potential for cell-based repair strategies for various myocardial disorders[3]. Here we report the identification in mouse of a previously unknown cardiac myocyte lineage that derives from the proepicardial organ. These progenitor cells, which express the T-box transcription factor Tbx18, migrate onto the outer cardiac surface to form the epicardium, and then make a substantial contribution to myocytes in the ventricular septum and the atrial and ventricular walls. Tbx18-expressing cardiac progenitors also give rise to cardiac fibroblasts and coronary smooth muscle cells. The pluripotency of Tbx18 proepicardial cells provides a theoretical framework for applying these progenitors to effect cardiac repair and regeneration.

Emergence of recent data has generated a paradigm shift for our understanding of cardiogenesis, with consequent implications for an understanding of cardiac progenitors and the aetiology of congenital heart disease. It was recognized that cardiac muscle cells derived from precardiac mesoderm subsequently form the primitive heart tube. More recently, the discovery of the secondary or anterior heart field, which contributed cells to the outflow tract and potentially to the right ventricle, suggested the presence of two distinct cardiac lineages[4]. Subsequent lineage studies based on expression of the LIM homeodomain transcription factor *Islet 1* (*Isl1*), and retrospective clonal analysis in the mouse, have confirmed two cardiac lineages, the first and second lineage, based on their timing of entry into the heart and the timing of their differentiation[4–6]. Here we report a previously unknown myocardial lineage, derived from *Tbx18*-expressing epicardial cells, that makes a substantial contribution to the heart.

Previous studies in avian species have demonstrated that the proepicardium and/or epicardium is a source for coronary vascular progenitors and cardiac fibroblasts[7–9]. During embryogenesis, cells from the proepicardium emigrate to form the epicardium—the epithelial outer lining of the heart. Epicardial cells undergo an epithelial-to-mesenchymal transition and invade the heart, giving rise to vascular endothelial cells, coronary vascular support cells and adventitial fibroblasts[7–9]. Recently, regeneration of the zebrafish heart has been shown to be associated with re-activation of an early marker of the proepicardium/epicardium, the T-box transcription factor *Tbx18* (refs 10, 11). *Tbx18*-expressing cells appear to cluster at the wound site of zebrafish heart, with concurrent appearance of neo-vasculature during regeneration[10].

To facilitate visualization of *Tbx18* expression, we generated an *nlacZ* (nuclear *lacZ*) knock-in into the endogenous *Tbx18* locus in mouse (Supplementary Fig. 3), and found that *nlacZ* expression mirrored that of the endogenous *Tbx18* gene (Fig. 1 and Supplementary Fig. 2). Neither *Tbx18* mRNA nor *Tbx18:nlacZ* expression are observed within the heart up to embryonic day (E)11.5 (Fig. 1).

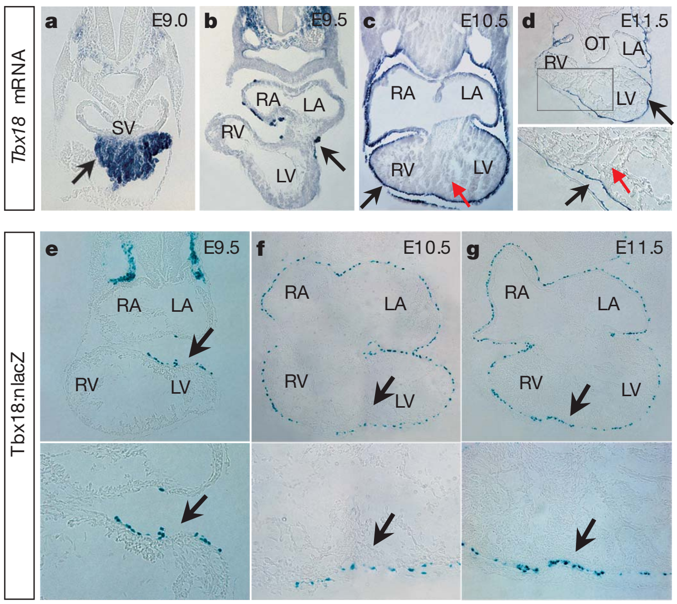


Figure 1 | LacZ expression in Tbx18:nlacZ knock-in mice recapitulates endogenous *Tbx18* expression. a–d, Section RNA *in situ* hybridization (ISH) of *Tbx18* in mouse embryos (E9.0–E11.5). *Tbx18* is expressed very early in the proepicardium (**a**) and epicardium (**b–d**). For parallel whole-mount RNA ISH, see Supplementary Fig. 2. **e–g**, X-gal staining on cardiac sections from Tbx18:nlacZ embryos at E9.5–E11.5. Tbx18:nlacZ cells are detected in the early epicardium at E9.5 (**e**) and in all epicardial cells covering the heart after E10.5 (**f, g**, data not shown for E12.5–E13.5). X-gal staining on Tbx18:nlacZ mouse tissues is consistent with *Tbx18* mRNA ISH. For the Tbx18:nlacZ targeting strategy, see Supplementary Fig. 3. Black arrows indicate *Tbx18* mRNA expression (**a–d**) and Tbx18:nlacZ cells (**e–g**); red arrows in **c** and **d** indicate that *Tbx18* is not expressed in heart between E9.5 and E11.5. For **e–g**, lower panels are high-magnification views of the upper panels in the heart (black arrow regions). LA/RA, left/right atrium; LV/RV, left/right ventricle; OT, outflow tract; and SV, sinus venosus.

[1]Skaggs School of Pharmacy, [2]Department of Bioengineering, [3]Department of Medicine, University of California, San Diego, La Jolla, California 92093, USA. [4]Department of Developmental and Regenerative Biology, [5]Centre for Molecular Cardiology, [6]The Black Family Stem Cell Institute, Mount Sinai School of Medicine, 1 Gustave L. Levy Place, New York, New York 10029, USA. [7]Complex Carbohydrate Research Centre, University of Georgia, Athens, Georgia 30602, USA. [8]Developmental Neurobiology Programme, The Burnham Institute, La Jolla, California 92037, USA. [9]Centre for Rheumatology, Royal Free and University College Medical School, Rowland Hill Street, London NW3 2PF, UK.
*These authors contributed equally to this work.

To investigate epicardial lineages in the mouse, we also generated a Cre knock-in into the endogenous *Tbx18* locus (Supplementary Fig. 5), and crossed Tbx18:Cre mice with the lineage reporter R26R^{lacZ} (ref. 12) mice. Analysis of *lacZ* expression in embryos from this cross demonstrated early expression consistent with that of endogenous *Tbx18* (ref. 11; Figs 1 and 2 and Supplementary Figs 2 and 7a, b). In contrast to active expression of *Tbx18*, lineage analysis revealed the presence of *Tbx18*-derived lineages within the heart by E10.5 in the region of the forming ventricular septum and in scattered regions within both ventricular walls and atria (Fig. 2d–f and Supplementary Fig. 7b). Co-immunostaining with cardiac troponin T (*cTnT*, also known as *cTnnt*), cardiac troponin I (*cTnI*, also known as *cTnni*), *MF20* (a sarcomeric myosin antibody) and the transcription factors *Gata4* and *Nkx2.5* demonstrated that these *Tbx18*-derived cells were cardiomyocytes (Fig. 2g–i and Supplementary Fig. 6). *Tbx18*-lineage-traced, *Nkx2.5*-positive cells could first be observed within the heart at E9.75 (Supplementary Fig. 6a). Complementary lineage studies using an organ explant culture system were consistent with Tbx18:Cre/R26R^{lacZ} lineage studies. Outer epicardial cells of embryonic hearts (E11–E13) were selectively labelled with 5-(and 6-)carboxyfluorescein diacetate succinimidyl ester (CFSE)[13]. Hearts harvested at time 0 exhibited specific labelling of the epicardium (Supplementary Fig. 7c). After culture for 18–24 h, fluorescently labelled epicardial cells were observed within the heart, and expressed *cTnT* (Supplementary Fig. 7d–f).

In contrast to earlier stages of heart development, cells actively expressing *Tbx18* are present within the heart at E12.5 (ref. 14). We examined sections from Tbx18:nlacZ mice from E10.5 to E17.5 and from neonatal stages using X-gal staining to detect β-galactosidase enzyme, encoded by the *lacZ* gene. (Fig. 1f, g and Supplementary Fig. 4d–f, data not shown for other stages), and by co-immunostaining with β-galactosidase antibody and cell-type-specific markers including *cTnT* and *Nkx2.5* (myocytes), *Pdgfrb* (vascular support cells[15]) and *Pecam1* (endothelial cells). Results demonstrated that cells actively expressing *Tbx18* within the heart at or after E12.5 were neither myocytes (*Nkx2.5*-negative, Supplementary Fig. 4a–c, data not shown for other stages) nor endothelial cells (*Pecam1*-negative, data not shown). Using mice containing a fibroblast-specific promoter from the *Col1a2* gene driving expression of a fusion protein in which cre recombinase is fused to a tamoxifen-inducible mutant oestrogen receptor (ER(T))[16], and the *Tbx18*-floxed nlacZ/nGFP (nuclear green fluorescent protein) allele (Supplementary Fig. 3), we demonstrated that some *Tbx18*-expressing cells were fibroblasts (Supplementary Fig. 4g). Also, a population of Tbx18:nlacZ cells co-expressed *Pdgfrb* (Supplementary Fig. 4h panels 1–6), a marker of vascular support cells (both pericytes and vascular smooth muscle)[15]. *Tbx18*-expressing fibroblasts were not stained with the Pdgfrb antibody (Supplementary Fig. 4i, panels 1–6), demonstrating that both fibroblasts and vascular support cells actively express *Tbx18* within the heart. Active expression of *Tbx18* was not observed in cardiomyocytes, except within the sinus horns[17].

Adult lineage analysis revealed that *Tbx18* derivatives were observed in the differentiated smooth muscle of the coronary vessels (Fig. 3a, b, c (insets 2 and 3), h, green or white arrows), including the coronary arteries, as evidenced by co-localization of β-galactosidase antibody with neuropilin-1 (*NP-1*, also known as *Nrp1*; Fig. 3i), a specific marker for arteries[18]. *Tbx18* expression was maintained in some coronary vascular smooth muscle cells (Fig. 3j, k). At adult stages, a substantial population of ventricular and atrial myocytes were observed to derive from *Tbx18*-expressing progenitors (Fig. 3c–f and Supplementary Fig. 8, black or white notched arrows). Cardiomyocytes isolated from Tbx18:Cre/R26R^{lacZ}-lineage-traced hearts exhibited X-gal-positive staining (Fig. 3g). Calcium transient analysis performed on myocytes isolated from Tbx18:Cre/R26R^{EYFP}-lineage-traced adult hearts demonstrated no differences between *Tbx18*-labelled myocytes and their non-labelled counterparts (Supplementary Fig. 9). *Tbx18* descendant cells also contributed significantly to the atrioventricular valves (Fig. 3c, inset 4). Co-localization of *Tbx18*-derived cells with endothelial lineages (as marked by *Pecam1*) was not observed in either embryonic or at adult stages (Fig. 3l).

Because previous studies have demonstrated that epicardial cells give rise to coronary endothelial cells[7–9,19], we verified immunohistochemical results by X-gal staining of affinity-purified endothelial cells isolated from *Tbx18*-lineage-traced hearts (Fig. 3m–o). Purified cardiac endothelial cell fractions did not contain *Tbx18*-lineage-positive cells (Fig. 3o, right panel). We then investigated whether the proepicardium contained endothelial lineages distinct from *Tbx18*-expressing cells. Immunostaining demonstrated a population of *Flk-1* (also known as *Kdr*) cells within the proepicardium that did not express *Tbx18* (Fig. 4a). No *Nkx2.5*-positive myocardial progenitors were observed within the proepicardium (Fig. 4b). We also isolated fibroblasts from *Tbx18*-lineage-traced hearts. Consistent with our Col1a2-Cre-ER(T)[16] results (Supplementary Fig. 4g), approximately one-third of cardiac fibroblasts expressed β-galactosidase (Fig. 3p).

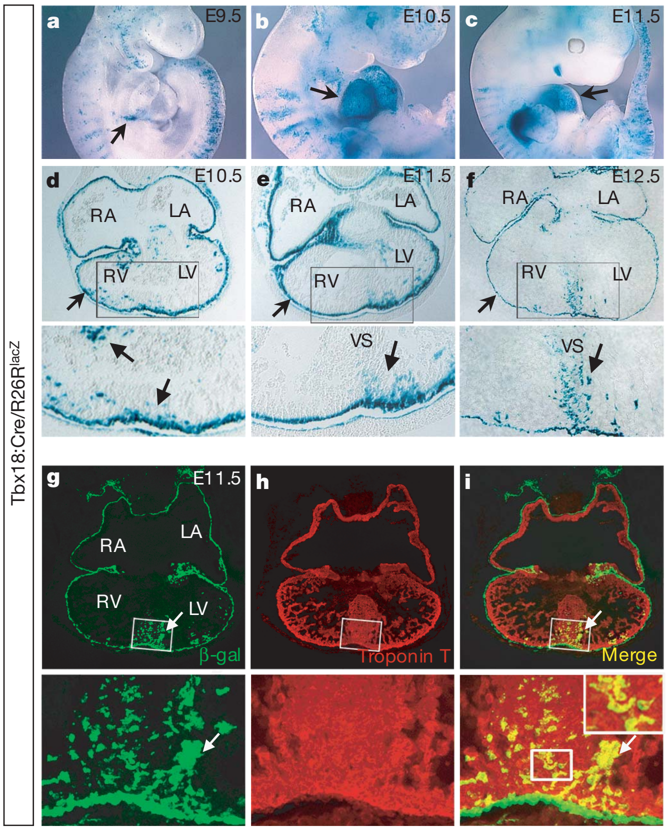


Figure 2 | Cells derived from *Tbx18*-expressing cells are observed within the heart by E10.5, and exhibit a cardiomyocyte identity. Tbx18:Cre mice were crossed to R26R^{lacZ} indicator mice[12] (Tbx18:Cre/R26R^{lacZ}). **a**–**c**, X-gal staining on whole-mount embryos (E9.5–E11.5) shows *Tbx18*-lineage-traced cells in the proepicardium and epicardium (arrows in **a**–**c**). **d**–**f**, X-gal staining on cryosections from E10.5 to E12.5. *Tbx18*-lineage-traced cells are observed in the epicardium (arrows in the upper panels of **d**–**f**) and within the heart, particularly within the apical region in the developing ventricular septum (arrows in the lower panels of **d**–**f**). The lower panels of **d**–**f** show high-magnification views of the ventricles in the upper panel micrographs. For *Tbx18*-lineage-traced cells at E13.5, see Supplementary Fig. 7. **g**, β-galactosidase antibody staining on *Tbx18*-lineage-traced tissue at E11.5. **h**, Cardiac troponin T antibody staining. **i**, Overlay of **g** and **h** revealed that *Tbx18*-lineage-traced cells in heart were cardiac-troponin-T-positive cells. In **g**–**i**, lower panels are high-magnification views of the upper panels in the ventricular septum region. Arrows in **d**–**i** indicate *Tbx18*-lineage-traced cells in the ventricular septum region. LA/RA, left/right atrium; LV/ RV, left/right ventricle; and VS, ventricular septum.

Isl1-expressing second heart field progenitors contribute cardiomyocytes that ultimately reside in the outflow tract, the right ventricle, the ventricular septum, the left ventricle, the atria and the atrial septum[5,20]. *Tbx18*-derived cardiomyocytes contributed significantly to the ventricular septum and atria as well as to small numbers of cells scattered throughout both ventricular walls (Figs 2d–i and 3c–g and Supplementary Figs 7a, b and 8). A comparison of Tbx18:Cre, Isl1:Cre- and Isl1:Cre/Tbx18:Cre-lineage-traced hearts indicated complementarity of these two lineages (Fig. 3c and Supplementary Figs 10 and 11a, b), as demonstrated by quantitative analysis of myocytes isolated from lineage-traced hearts (ventricular septum, left ventricle and right ventricle; Supplementary Table 1). Immunostaining and RNA *in situ* hybridization of *Tbx18*, *Isl1* and *MLC2a* (also known as *Myl7*) revealed that *Tbx18* was not co-expressed with *Isl1* and *MLC2a* during early embryogenesis (Supplementary Figs 2a–e and 11c).

We isolated Tbx18:Cre/R26R^{EYFP}-lineage-traced proepicardial cells and cultured them under conditions to favour differentiation into cardiomyocytes or smooth muscle cells[21]. Immunostaining analyses demonstrated efficient conversion of *Tbx18* lineages to myocytes or smooth muscle cells (Supplementary Fig. 12). To evaluate the pluripotency of individual proepicardial cells, single-cell clonal analysis was performed. Out of 336 single proepicardial cells plated on OP9 feeder layers, approximately 37% (124 out of 336) proliferated and formed clones by day seven. Forty clones were picked randomly, each dispersed into two wells, and cultured under conditions to favour either myocyte or smooth muscle cell fates. After five days of culture in differentiation medium, 34% of single-cell clonal derivatives differentiated into cardiomyocytes with an obvious striated cytoarchitecture and expressed *cTnT* (Fig. 4c). Spontaneous contraction was observed in some wells after four days of culture (Supplementary Fig. 13 and Supplementary Video), and myocyte identity was further confirmed by calcium transients (Fig. 4e–g). Each clone (40 out of 40) cultured with smooth muscle culture medium stained with smooth muscle myosin heavy chain (Fig. 4d). No instances were observed where derivatives of a single clone formed only cardiomyocytes and not smooth muscle cells. This demonstrated that a large proportion of proepicardial cells are pluripotent and can adopt either cardiomyocyte or smooth muscle cell fates.

Adult epicardial cells can be activated to migrate *in vitro* and adopt vascular cell fates[22]. Migratory adult epicardial cells from Tbx18:nlacZ mice expressed *Tbx18* (Supplementary Fig. 14). To

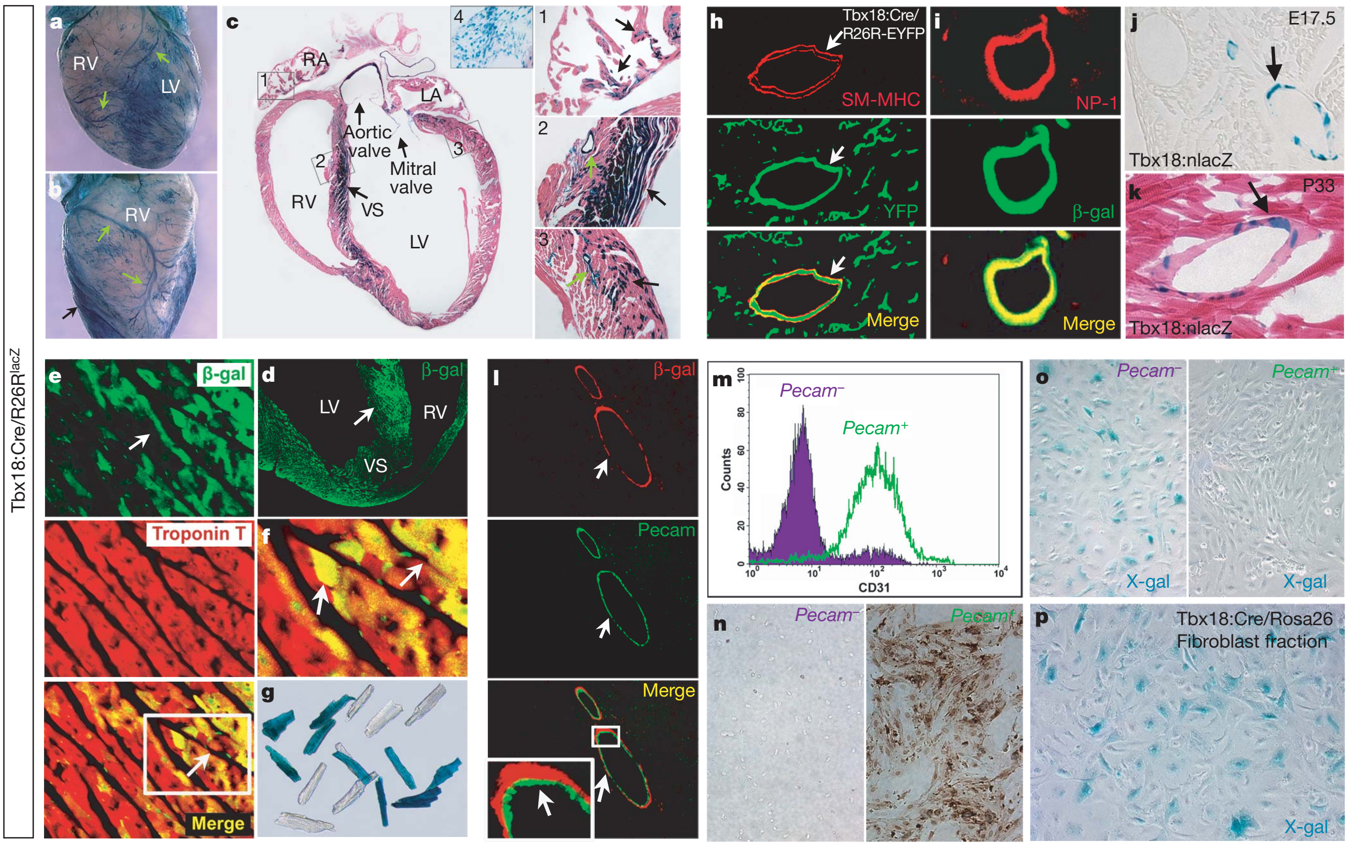


Figure 3 | *Tbx18* lineage tracing in the adult heart. **a**, **b**, Whole-mount X-gal staining of Tbx18:Cre/R26R^{lacZ} adult mouse heart (6 weeks). Coronary vasculature (green arrows) is derived from *Tbx18* lineages. Dense staining in the septum is visible (black arrow) **c**, X-gal staining of tissues from Tbx18:Cre/R26R^{lacZ} adult mouse heart (6 weeks). *Tbx18* can give rise to: cardiomyocytes within atria (arrows in **c**, inset 1 and Supplementary Fig. 8), the ventricular septum (black arrow in **c**, inset 2) and the ventricular wall (black arrow in **c**, inset 3); coronary vascular support cells (green arrows in **c**, insets 2 and 3); and atrioventricular valves (**c**, inset 4, high magnification for bicuspid valve). Most *Tbx18*-lineage-traced cells within the ventricular septum co-stained with cardiac troponin T (arrows in **d**–**f**; **f** is a high magnification of the lower panel of **e**), demonstrating they are cardiomyocytes. **g**, A subset of cardiomyocytes isolated from adult *Tbx18*-lineage-traced hearts exhibited X-gal-positive staining. **h**, *Tbx18* lineages give rise to coronary vascular smooth muscle cells (co-localized with smooth muscle myosin heavy chain). **i**, *Tbx18* lineages give rise to coronary artery smooth muscle cells (co-localized with *NP-1*, an artery marker). **j**, **k**, X-gal staining on Tbx18:nLacZ knock-in mice shows that *Tbx18* expression is maintained in some coronary vascular smooth muscle cells from embryonic stages to adulthood. **l**–**o**, *Tbx18* lineages do not give rise to coronary vascular endothelial cells (not co-localized with *Pecam1*, **l**). **m**, Coronary vascular endothelial cells were isolated from the hearts (green fraction)[26] and react with Pecam1 antibody (**n**, right panel), whereas *Pecam1*$^{-}$ fraction (purple) cells do not (**n**, left panel). **o**, Purified *Pecam1*$^{+}$ cells are Xgal^{-} (**o**, right panel), confirming that *Tbx18*-expressing lineages do not give rise to coronary vascular endothelial cells. A portion of cells in *Pecam1*$^{-}$ fraction are Xgal^{+} (**o**, left panel). **p**, Cardiac fibroblasts isolated from *Tbx18*-lineage-traced hearts[25] demonstrated that approximately 30% of these cells derive from *Tbx18* lineages. Abbreviations are as in Fig. 2.

investigate whether postnatal and adult epicardium retains a similar pluripotency to that of the proepicardium, postnatal and adult epicardial cells isolated from Tbx18:Cre/R26R^{EYFP}-lineage-traced mice were cultured on OP9 cells, and then subsequently cultured in myocyte- or smooth-muscle-specific differentiation medium. In contrast to results with proepicardial cells, adult epicardial cells did not convert to *cTnT*-expressing cells under these culture conditions (data not shown). Understanding the underlying causes of this difference in potential will have significant implications for the possible use of epicardial cells for cardiac repair.

Our data are consistent with a model whereby the first cells to enter the heart from the proepicardium/epicardium give rise to myocyte lineages, first observed at approximately E9.75, whereas subsequent *Tbx18*-positive epicardial lineages that will give rise to vascular support cells and fibroblasts are first observed entering the heart at approximately E12.5. Endothelial lineages within the proepicardium are distinct from those of *Tbx18* cells (Supplementary Fig. 1). *Tbx18* is not actively expressed within cardiac myocyte derivatives, but is observed in a subset of vascular smooth muscle and cardiac fibroblasts, in addition to the epicardium (Fig. 3j, k and Supplementary Fig. 4g–i).

Previous studies in avian embryos have failed to demonstrate that proepicardial lineages give rise to cardiomyocytes within the heart, although myocytic potential has been demonstrated *in vitro*[7–9,23]. It is unknown whether observed differences reflect distinct experimental approaches or species-specific differences. In mouse, *Tbx18*-expressing cells of the septum transversum that are in continuity with the sinus venosus contribute to the myocardial sleeve of the latter[17]. These cells, however, never express *Nkx2.5*, in contrast to the *Tbx18*-derived intracardiac myocytes of the ventricles and atria described here. The distinct embryonic origin of a substantial number of cardiomyocytes from an epicardial lineage provides a new perspective on heart development and congenital or adult heart disease affecting these lineages.

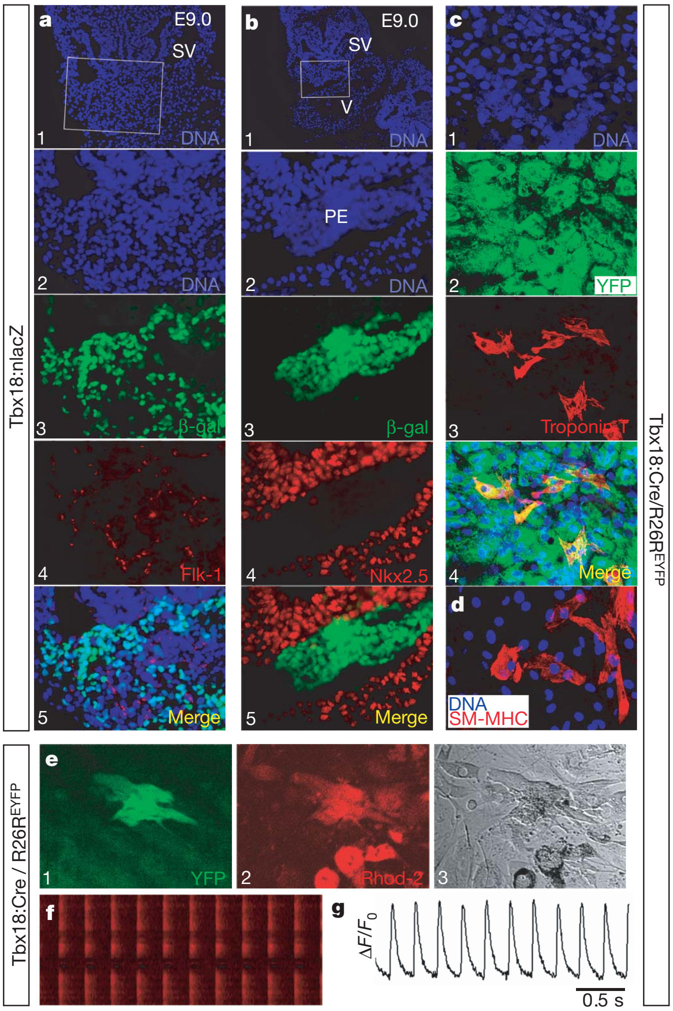


Figure 4 | *Tbx18*-expressing progenitors within proepicardium are distinct, and retain the capacity to differentiate into cardiomyocytes and smooth muscle cells *in vitro*. **a**, **b**, *Tbx18*-expressing progenitors within the proepicardium are not co-stained with antibodies to Flk-1 (**a**) or Nkx2.5 (**b**) (E9.0, data not shown for E9.5), suggesting that *Flk-1* endothelial progenitors within the proepicardium are a distinct population from *Tbx18*-expressing cells. *Tbx18* progenitors with myocyte potential do not yet express *Nkx2.5* within the proepicardium, although *Tbx18*-derived myocytes do express *Nkx2.5* within the heart. The second panels in **a** and **b** are high-magnification images of the top panels of **a** and **b**, respectively. **c**, *Tbx18*-lineage-traced proepicardial cells (Tbx18:Cre/R26R^{EYFP}) were disassociated and single cells were expanded in proepicardial culture medium. A significant percentage of clones (34%) derived from single proepicardial cells adopt cardiomyocyte fate after incubating in the differentiation medium[21] (co-stained with cardiac troponin T, bottom panel of **c**). **d**, Every clone (40 out of 40) derived from single *Tbx18* proepicardial cells displayed staining for smooth muscle myosin heavy chain (SM-MHC) antibody after incubating in smooth muscle culture medium. **e**–**g**, Real-time Ca^{2+} transients of cells derived from single *Tbx18*-lineage-traced proepicardial cells after myocytic differentiation. **e**, Left panel: monitored cell is YFP$^+$, suggesting it derived from the Tbx18:Cre/R26R^{EYFP} proepicardium. **e**, Middle panel: differentiated cells were monitored under an inverted confocal microscope after they were preloaded with Ca^{2+} indicator Rhod-2-AM. **e**, Right panel, transmitted light image of the monitored cell. **f**, Line-scanning images and time course (**g**) of spontaneous Ca^{2+} transients obtained from differentiated YFP$^+$ cells. PE, proepicardium; SV, sinus venosus; and V, ventricle.

METHODS SUMMARY

Whole-mount RNA *in situ* hybridization was performed on mouse embryos as described previously[24]. Tbx18:nlacZ/nGFP and Tbx18:Cre knock-in mouse models were generated to visualize *Tbx18*-expressing and lineage-traced cells throughout heart development, respectively. Fate-mapping of E11.5 epicardial cells was achieved by *in situ* dye labelling followed by heart explant culture *in vitro*[13]. Pure populations of adult cardiomyocytes and fibroblasts were isolated by means of perfused collagenase digestion by canulation of the aorta[25]. Adult heart endothelial cells were affinity-purified by Pecam1 antibody[26]. Adult epicardium was manually peeled as a monolayer sheet after a brief collagenase treatment. Pure isolations of *Tbx18*-lineage-traced proepicardial cells were accomplished by localized suction with a pulled-tip glass pipette. For single-cell clonal analysis, trypsin was used to dissociate cells. Single proepicardial cells were plated onto inactivated OP9 feeder layers. After proliferation, clones were dispersed and preplated onto gelatin, where differentiation was induced with specially designed media conditions[21]. Line-scan imaging was used to measure Ca^{2+} transients of myocytes loaded with Rhod-2-AM[27]. All immunohistochemical experiments were performed on cryosections of 4% paraformaldehyde-fixed tissues or cultured cells to determine lineage specification.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 December 2007; accepted 1 April 2008.
Published online 14 May 2008; corrected 3 July 2008 (details online).

1. Olson, E. N. A decade of discoveries in cardiac biology. *Nature Med.* 10, 467–474 (2004).
2. Srivastava, D. Making or breaking the heart: from lineage determination to morphogenesis. *Cell* 126, 1037–1048 (2006).
3. Murry, C. E., Field, L. J. & Menasche, P. Cell-based cardiac repair: reflections at the 10-year point. *Circulation* 112, 3174–3183 (2005).
4. Buckingham, M., Meilhac, S. & Zaffran, S. Building the mammalian heart from two sources of myocardial cells. *Nature Rev. Genet.* 6, 826–835 (2005).
5. Cai, C. L. *et al.* Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev. Cell* 5, 877–889 (2003).
6. Kelly, R. & Evans, S. M. The secondary/anterior heart field. *Heart Development and Regeneration* (eds Rosenthal, N. & Harvey R. P.) (Academic, San Diego, in the press).
7. Mikawa, T. & Fischman, D. A. Retroviral analysis of cardiac morphogenesis: discontinuous formation of coronary vessels. *Proc. Natl Acad. Sci. USA* 89, 9504–9508 (1992).

8. Mikawa, T. & Gourdie, R. G. Pericardial mesoderm generates a population of coronary smooth muscle cells migrating into the heart along with ingrowth of the epicardial organ. *Dev. Biol.* **174,** 221–232 (1996).
9. Gittenberger-de Groot, A. C., Vrancken Peeters, M. P., Mentink, M. M., Gourdie, R. G. & Poelmann, R. E. Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions. *Circ. Res.* **82,** 1043–1052 (1998).
10. Lepilina, A. *et al.* A dynamic epicardial injury response supports progenitor cell activity during zebrafish heart regeneration. *Cell* **127,** 607–619 (2006).
11. Kraus, F., Haenig, B. & Kispert, A. Cloning and expression analysis of the mouse T-box gene *Tbx18. Mech. Dev.* **100,** 83–86 (2001).
12. Soriano, P. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21,** 70–71 (1999).
13. Wilm, B., Ipenberg, A., Hastie, N. D., Burch, J. B. & Bader, D. M. The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature. *Development* **132,** 5317–5328 (2005).
14. Franco, D. *et al.* Left and right ventricular contributions to the formation of the interventricular septum in the mouse heart. *Dev. Biol.* **294,** 366–375 (2006).
15. Armulik, A., Abramsson, A. & Betsholtz, C. Endothelial/pericyte interactions. *Circ. Res.* **97,** 512–523 (2005).
16. Zheng, B., Zhang, Z., Black, C. M., de Crombrugghe, B. & Denton, C. P. Ligand-dependent genetic recombination in fibroblasts: a potentially powerful technique for investigating gene function in fibrosis. *Am. J. Pathol.* **160,** 1609–1617 (2002).
17. Christoffels, V. M. *et al.* Formation of the venous pole of the heart from an Nkx2–5-negative precursor population requires Tbx18. *Circ. Res.* **98,** 1555–1563 (2006).
18. You, L. R. *et al.* Suppression of Notch signalling by the COUP-TFII transcription factor regulates vein identity. *Nature* **435,** 98–104 (2005).
19. Perez-Pomares, J. M. *et al.* Origin of coronary endothelial cells from epicardial mesothelium in avian embryos. *Int. J. Dev. Biol.* **46,** 1005–1013 (2002).
20. Verzi, M. P., McCulley, D. J., De Val, S., Dodou, E. & Black, B. L. The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev. Biol.* **287,** 134–145 (2005).
21. Moretti, A. *et al.* Multipotent embryonic Isl1^{+} progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127,** 1151–1165 (2006).
22. Smart, N. *et al.* Thymosin β4 induces adult epicardial progenitor mobilization and neovascularization. *Nature* **445,** 177–182 (2007).
23. Kruithof, B. P. *et al.* BMP and FGF regulate the differentiation of multipotential pericardial mesoderm into the myocardial or epicardial lineage. *Dev. Biol.* **295,** 507–522 (2006).
24. Cai, C. L. *et al.* T-box genes coordinate regional rates of proliferation and regional specification during cardiogenesis. *Development* **132,** 2475–2487 (2005).
25. Subramanian, S. V., Kelm, R. J., Polikandriotis, J. A., Orosz, C. G. & Strauch, A. R. Reprogramming of vascular smooth muscle α-actin gene expression as an early indicator of dysfunctional remodeling following heart transplant. *Cardiovasc. Res.* **54,** 539–548 (2002).
26. Wang, L., Fuster, M., Sriramarao, P. & Esko, J. D. Endothelial heparan sulfate deficiency impairs L-selectin- and chemokine-mediated neutrophil trafficking during inflammatory responses. *Nature Immunol.* **6,** 902–910 (2005).
27. Ouyang, K. *et al.* Ca^{2+} sparks and secretion in dorsal root ganglion neurons. *Proc. Natl Acad. Sci. USA* **102,** 12259–12264 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Kleckner, J. Lam and M. Zamora for critical technical assistance; D. Bader for discussion; P. Soriano and F. Costantini for providing $R26R^{lacZ}$ and $R26R^{EYFP}$ indicator mice; H. Kubo for providing OP9 cells; and J. Lin, D. Fischman and A. Kolodkin for providing troponin I, Pdgfrb and neuropilin 1 antibodies, respectively. We are also grateful to B. Gelb for his critical comments and revisions of the manuscript. This work was supported by AHA National Scientist Development Grant to C.-L.C. and NIH1RO1 to S.M.E.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.M.E. (syevans@ucsd.edu).

METHODS

Whole-mount RNA *in situ* hybridization and histological analysis. Whole-mount RNA ISH was carried out as described previously[24].

Generation of Tbx18:nlacZ/nGFP, Tbx18:Cre and Isl1:Cre knock-in mouse models[28]. *Tbx18* and *Isl1* genomic clones were isolated by screening a mouse genomic lambda library (129/sv, Stratagene). Arms of targeting constructs were amplified on phage DNA templates with high-fidelity DNA polymerase (Pfu, Stratagene, catalogue number 600153). For individual targeting strategy, see Supplementary Figs 3, 5 and 10.

X-gal staining. For the whole-mount staining, mouse embryos and tissues were collected from timed pregnant females. Embryos and tisssues were fixed in 4% paraformaldehyde for 15–60 min. After permeabilization (10% Na-deoxycholate, 10% NP40 in PBS), embryos and tissues were stained in X-gal solution (50 mM K-ferricyanide, 50 mM K-ferrocyanide, 200 mM $MgCl_2$, 100 mg ml^{-1} X-gal in PBS) for 4–12 h and were then post-fixed with 4% paraformaldehyde. For section staining, embryos and tissues were fixed in 4% paraformaldehyde and were then dehydrated in series sucrose solution. X-gal staining was performed on 6–8-μM cryosections with previous additional fixation in 4% paraformaldehyde for 6–8 min.

Fluorescent dye lineage tracing of epicardial cells with a heart explant culture system. Embryos were dissected at E11.5 in PBS. About 20 μl dye solution (30 μg ml^{-1}) was injected into the pericardial cavity (CFSE, Molecular Probes, C1157). The chest expands under the pressure and it was ensured that it did not shrink back, indicating the dye was not able to leak from the cavity. Embryos were then submerged in media (DMEM) and left at room temperature (23 °C) for 15 min in the dark. Hearts were removed and washed in media for 5 min. Some hearts were selected randomly for fixation in 4% paraformaldehyde. The rest were cultured on a 0.4 μm polycarbonate membrane insert with 10% embryonic stem cell qualified fetal bovine serum (FBS) (Gibco, 10439024) supplemented DMEM (Gibco, D5796) for 18–24 h. Hearts were fixed and frozen as described previously for immunohistochemistry.

Isolation of cardiomyocytes, endothelial cells, fibroblasts and epicardial cells from adult mouse hearts. For cardiomyocyte isolation, pure adult cardiomyocytes (6–8 weeks) were isolated according to a previously described method[29]. For Ca^{2+} imaging and *Tbx18/Isl1* lineage contribution quantification, different parts of the heart (ventricular septum, left ventricular wall and right ventricular wall) were dissected and separated after retrograde collagenase perfusion via the aorta. After an additional collagenase incubation for 10 min at 37 °C, cardiomyocytes were dispersed mechanically[29]. Pure populations of myocytes were obtained by centrifugation and cell filtration. Quantification was performed on a number of independent hearts for each sample by cell counts of numerous, random, field-of-view micrographs to determine the percentage of total cardiomyocytes expressing the YFP lineage marker. Statistical analyses were performed on the data sets. For isolation of endothelial cells, adult mouse hearts were minced with a razor blade. Endothelial cells from the hearts were isolated according to a previously described method[26]. Cardiac fibroblasts were isolated and cultured as described previously[25]. Pre-plating of cells isolated from the heart digest was shortened to 30 min at 37 °C to increase the purity of fibroblasts. For epicardial cell isolation, adult epicardial cells were isolated from explanted hearts by treatment with collagenase at 37 °C for 10 min (prepared as described for cardiomyocyte isolation). The epicardium was then manually extracted as a monolayer sheet.

Immunostaining. Embryos and tissues were fixed immediately in 4% paraformaldehyde for 10–30 min after dissection. Tissues were embedded and cut by cryo-sectioning (5–10 μm). For sections with endogenous YFP, a post-fixation for 5 min on ice was performed before staining. Cells obtained in culture were washed with warm media and then were fixed for 7 min with 10% formalin. Primary antibodies used in this study were: rabbit polyclonal anti-β-galactosidase (Cappel, product number 55978, 1:200), goat polyclonal anti-β-galactosidase (Biogenesis, 4600–1409, 1:200), rabbit polyclonal anti-smooth muscle myosin heavy chain (Biomedical Technologies Inc., BT562, 1:200), rabbit polyclonal anti-NKX2.5 (Santa Cruz Biotechnology, SC14033, 1:50), rat polyclonal anti-Pecam1 (Pharmingen, 550274, 1:100), mouse monoclonal anti-cardiac troponin T (NeoMarkers, MSZ-295-P, 1:200), mouse monoclonal anti-α-smooth muscle actin (Abcam, ab7817, 1:200), mouse monoclonal anti-α-actinin (sarcomeric) (Sigma-Aldrich, A7811, 1:200), rabbit polyclonal anti-WT1 (Santa Cruz Biotechnology, sc-846, 1:75), goat polyclonal anti-GATA-4 (Santa Cruz Biotechnology, sc-1237, 1:75), mouse polyclonal anti-Isl1 (Developmental Studies Hybridoma Bank, 39.4D5, 1:100). Mouse monoclonal cardiac troponin I was provided by J. Lin (Developmental Studies Hybridoma Bank, 1:100), mouse monoclonal anti-MF20 (1:200) was provided by D. Fischman, and rabbit polyclonal anti-Pdgfrb (1:200) was provided by W. Stallcup. Rabbit polyclonal anti-neuropilin 1 (1:200) was provided by A. L. Kolodkin.

Isolation and differentiation of proepicardial cells. *Tbx18*-lineage-traced embryos (Tbx18:Cre/$R26R^{EYFP}$, E9.0–E9.5) were dissected from timed pregnant mice and were quickly washed in cold PBS after dissection. A check was made under the epifluorescent microscope that *Tbx18*-positive cells in the proepicardium have not begun to migrate to the heart. Proepicardial cells were isolated with a pulled-tip glass pipette. The isolated proepicardial cells were cultured on gelatin with a medium designed to induce differentiation into cardiac myocytes or smooth muscle cells[21].

Single proepicardial cells preparation. Proepicardial cells collected from seven to ten E9.0–E9.5 embryos (Tbx18:Cre/$R26R^{EYFP}$) were digested with 0.05% trypsin/EDTA (Invitrogen) for 5 min at 37 °C. Single-cell suspension was obtained by pipetting cells in culture medium (85% IMDM (Gibco), 15% FBS (selected batches), 100 μM non-essential amino acids (Gibco), 2 mM sodium pyruvate (Gibco) and 100 μM β-mercaptoethanol (Sigma)). Disassociated cells were plated at 0.5 cells per well onto a 96-well plate with a mitomycin C (Sigma)-inactivated OP9 cell feeder layer (YFP signals were examined to assure single-cell colonies derived from Tbx18 lineages). After seven days of culture, clones proliferated from single cells were dispersed to small clumps with 0.05% trypsin/EDTA for 30 s. These cells were re-plated onto gelatin-coated 16-well chamber slides with one colony split into two wells and then left in stem cell culture medium for two days. Cells were cultured with a medium designed to induce differentiation into cardiac myocytes or smooth muscle cells[21]. Immunostaining was performed after five days of culture.

Ca^{2+} imaging. Cells were loaded with Rhod-2-AM (5 μM, 15 min; Molecular Probes, R1245MP), and imaged with an Olympus Fluoview 1000 inverted confocal microscope with a ×40 oil immersion lens (numerical aperture 1.3). The line-scan imaging mode was used to measure Ca^{2+} transients of paced cells. The extracellular solution contained 116 mM NaCl, 5.0 mM KCl, 0.8 mM $MgSO_4$, 1.0 mM NaH_2PO_4, 5.5 mM glucose, 20 mM HEPES and 1.0 mM $CaCl_2$ (pH 7.4). Image processing and data analysis were performed as previously described[27].

28. Hadjantonakis, A. K., Dickinson, M. E., Fraser, S. E. & Papaioannou, V. E. Technicolour transgenics: imaging tools for functional genomics in the mouse. *Nature Rev. Genet.* **4**, 613–625 (2003).
29. Zhou, Y. Y. *et al.* Culture and adenoviral infection of adult mouse cardiac myocytes: methods for cellular genetic physiology. *Am. J. Physiol. Heart Circ. Physiol.* **279**, H429–H436 (2000).

LETTERS

Epicardial progenitors contribute to the cardiomyocyte lineage in the developing heart

Bin Zhou[1,2], Qing Ma[1,2], Satish Rajagopal[1,2], Sean M. Wu[3], Ibrahim Domian[3], José Rivera-Feliciano[2], Dawei Jiang[1], Alexander von Gise[1,2,4], Sadakatsu Ikeda[1,2], Kenneth R. Chien[3] & William T. Pu[1,2]

The heart is formed from cardiogenic progenitors expressing the transcription factors Nkx2-5 and Isl1 (refs 1 and 2). These multipotent progenitors give rise to cardiomyocyte, smooth muscle and endothelial cells, the major lineages of the mature heart[3,4]. Here we identify a novel cardiogenic precursor marked by expression of the transcription factor Wt1 and located within the epicardium—an epithelial sheet overlying the heart. During normal murine heart development, a subset of these *Wt1*$^+$ precursors differentiated into fully functional cardiomyocytes. *Wt1*$^+$ proepicardial cells arose from progenitors that express Nkx2-5 and Isl1, suggesting that they share a developmental origin with multipotent *Nkx2-5*$^+$ and *Isl1*$^+$ progenitors. These results identify *Wt1*$^+$ epicardial cells as previously unrecognized cardiomyocyte progenitors, and lay the foundation for future efforts to harness the cardiogenic potential of these progenitors for cardiac regeneration and repair.

Epicardial cells migrate from the proepicardium, an outgrowth of the septum transversum, and spread over the surface of the heart[5,6]. A subset of epicardial cells transition to a mesenchymal phenotype, migrate into the subjacent myocardium and differentiate into smooth muscle and endothelial cells (Supplementary Fig. 1)[7–13]. Wt1 was expressed in proepicardium and epicardium, but not in myocardium (Fig. 1a–c). To trace the fate of *Wt1*-expressing proepicardial and epicardial cells, a *GFPCre* complementary DNA[14] was knocked into the endogenous *Wt1* start codon (Supplementary Fig. 2). GFP and Cre expression in *Wt1*$^{GFPCre/+}$ embryos co-localized with Wt1, indicating that the knock-in strategy placed *GFPCre* under the control of endogenous *Wt1* regulatory elements (Fig. 1d–f). In the heart, GFPCre expression was confined to proepicardium and epicardium from embryonic day (E) 9.5 to E15.5, and was not found in the myocardium (Fig. 1d–g).

We used *Wt1*^{GFPCre} and the Cre-activated reporters *Rosa26*^{fsLz} (ref. 15) and *Z/Red* (ref. 16) to analyse the fate of *Wt1*-expressing cells in the heart. On Cre-mediated recombination, these reporters heritably express β-galactosidase (β-gal) or red fluorescent protein (RFP), respectively. The use of two different reporters minimized potential artefacts related to unanticipated behaviour of Cre-activated reporters, or to false-positive immunostaining. Whereas Wt1 and GFPCre expression were confined to the epicardium, descendants of *Wt1*-expressing progenitors (hereafter called *Wt1*-derived cells), marked by β-gal, were found in a mosaic pattern throughout the myocardium (Fig. 2a). Consistent with previous reports showing that epicardially-derived mesenchyme predominately differentiates into smooth muscle cells in mammals, most *Wt1*-derived cells adopted a smooth muscle cell fate, and only a minority differentiated into endothelial cells (Supplementary Fig. 3)[8,9].

Notably, we found that some *Wt1*-derived cells differentiated into cardiomyocytes during normal heart development, as demonstrated

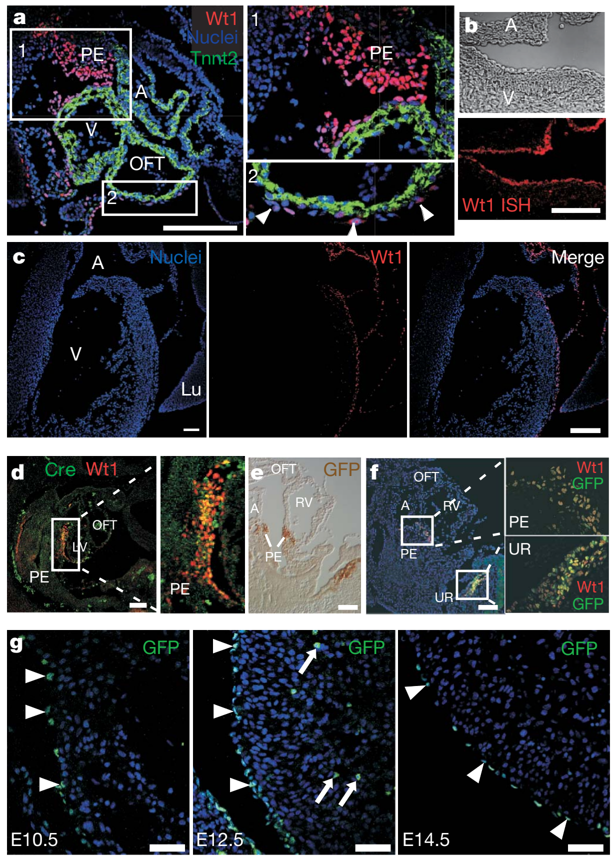


Figure 1 | Cardiac Wt1 and Wt1-driven GFPCre expression. **a**, At E9.5, Wt1 was expressed in proepicardium (PE) and in scattered pericardial cells over the surface of the heart (arrowheads in magnified image of inset 2). **b**, **c**, At E15.5, Wt1 expression was confined to the epicardium, as determined by *in situ* hybridization (ISH, **b**) and immunohistochemistry (**c**). **d**–**f**, Co-expression of Wt1 and GFPCre in *Wt1*$^{GFPCre/+}$ E9.5 embryos. Wt1 and GFPCre were co-expressed in proepicardium and urogenital ridge (UR). **g**, Wt1-driven GFPCre was confined to the epicardium (arrowheads) and not detected within the myocardium at E9.5–E15.5 (representative images for E10.5, E12.5 and E14.5 are shown). Arrows indicate autofluorescent red blood cells. All images except **b** show immunohistochemical staining. A, atrium; V, ventricle; LV and RV, right and left ventricle, respectively; Lu, lung; OFT, outflow tract; scale bars, 50 μm.

[1]Harvard Stem Cell Institute and Department of Cardiology, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, USA. [2]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA. [3]Harvard Stem Cell Institute, Harvard University and Cardiovascular Research Center, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [4]Clinic of Neonatology, Charité Campus Mitte, Charité Universitätsmedizin Berlin, Chariteplatz 1, 10117 Berlin, Germany.

by co-expression of lineage tracers and cardiomyocyte markers cardiac troponin T2 (Tnnt2) and sarcomeric-actinin (Actn1) (Fig. 2b–d). The *Wt1*-derived cardiomyocytes also expressed cardiac transcription factors Gata4 and Nkx2-5 (Supplementary Fig. 4). *Wt1*-derived cardiomyocytes were located in the myocardium of all four cardiac chambers and in the interventricular septum, constituting 7%–10% of cardiomyocytes in ventricles and 18% in atria (Supplementary Fig. 5). We further confirmed co-expression of cardiomyocyte and lineage tracers in isolated cells by immunostaining of dissociated fetal heart cultures (Fig. 2e–f). Four per cent of cardiomyocytes in dissociated fetal heart cultures were *Wt1*-derived (Supplementary Fig. 5), comparable to the frequency observed in tissue sections.

To determine if *Wt1*-derived cardiomyocytes had functional properties of cardiomyocytes, we analysed dissociated cells from $Wt1^{GFPCre/+}$;*Z/Red* fetal hearts. A subset of red fluorescent cells exhibited spontaneous contractile activity (Fig. 2g and Supplementary Movie 1). In addition, these contracting RFP^+ cells exhibited calcium oscillations, with kinetics, amplitude and frequency characteristic of cardiomyocytes (Fig. 2h–j and Supplementary Movie 2). Also characteristic of cardiomyocytes were calcium 'sparks' preceding calcium waves (Fig. 2h–1) and caffeine augmentation of calcium transient amplitude (Fig. 2j), consistent with calcium release from cardiomyocyte ryanodine receptors. Calcium transients of RFP^+ cells were indistinguishable from, and synchronous with, adjacent RFP^- cells (Fig. 2i), suggesting electrical coupling between *Wt1*-derived and non-*Wt1*-derived cardiomyocytes. Consistent with this finding, the gap junction protein connexin 43 (Cx43) localized to the membrane between *Wt1*-derived and neighbouring cardiomyocytes (Supplementary Fig. 6a). A similar pattern of Cx43 expression was observed in *Wt1*-derived cardiomyocytes in tissue sections of E15.5 hearts (Supplementary Fig. 6b), suggesting that *Wt1*-derived cardiomyocytes are also electrically coupled to other cardiomyocytes *in vivo*. Collectively, these data indicate that during heart development a subset of *Wt1*-expressing cells differentiate into cardiomyocytes.

To characterize the *Wt1*-expressing cardiomyocyte precursors further, we used several independent methods to control the temporal and spatial window during which these precursors were labelled. To regulate Cre-labelling temporally, we knocked a cDNA encoding a Cre-modified oestrogen receptor ligand-binding domain (*CreERT2*) into the *Wt1* locus (Supplementary Fig. 7). CreERT2 fusion protein recombinase activity requires tamoxifen[17]. Maternal injection of tamoxifen at E10.5 and E11.5 induced Cre activity and resulted in β-gal expression within $Wt1^{CreERT2}$;$Rosa26^{fsLz}$ myocardium (Fig. 3a–e), whereas $Wt1^{CreERT2}$ did not activate $Rosa26^{fsLz}$ in the absence of tamoxifen (Fig. 3b and Supplementary Fig. 7). The frequency of $Wt1^{CreERT2}$-labelled cells in epicardium and myocardium (Supplementary Figs 7 and 8) was reduced compared to constitutive labelling by $Wt1^{GFPCre}$, probably due to inefficient CreERT2

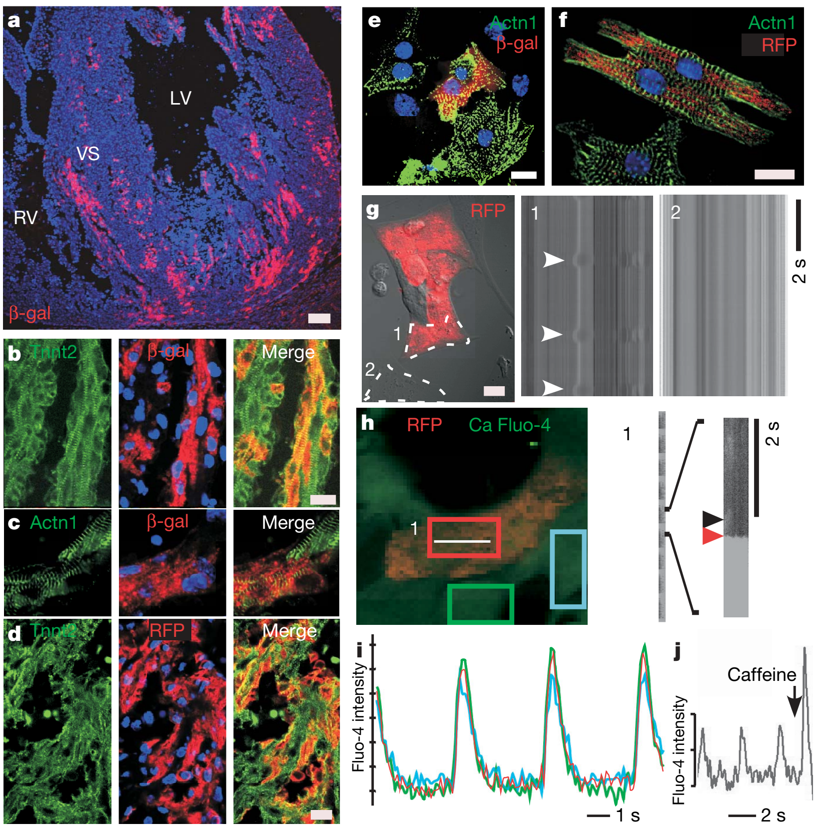


Figure 2 | *Wt1*-derived cells differentiate into cardiomyocytes. a–j, The fate of *Wt1*-derived cells, marked by β-gal expression in $Wt1^{GFPCre/+}$;$Rosa26^{fsLz}$ hearts (**a–c**, **e**), or RFP expression in $Wt1^{GFPCre/+}$;*Z/Red* hearts (**d**, **f**, **g–j**), was analysed by immunohistochemistry (**a–f**) and live cell imaging (**g–j**). **a**, Overview of the contribution of *Wt1*-derived cells to myocardium. **b–d**, Cryosections of E15.5 heart stained for genetic lineage tracers (β-gal or RFP) and cardiomyocyte-specific markers Tnnt2 and Actn1. **e–f**, Co-expression of genetic lineage tracers with cardiomyocyte markers in dissociated E15.5 heart culture. **g–j**, Live cell imaging of *Wt1*-derived cells, identified by RFP fluorescence. **g**, Spontaneous contraction of an RFP^+ cell. The transmission line scan along the paths is indicated by dashed lines. The arrowheads in inset 1 indicate contraction of an RFP^+ cell; a non-contractile cell is also shown (inset 2). **h**, Fluo-4 AM calcium imaging of a cluster of beating cells, one of which was RFP^+. Calcium oscillations in the RFP^+ cell, measured by line scan along the white line, showed calcium sparks (black arrowhead, inset 1) preceding calcium waves (red arrowhead). **i**, Calcium oscillations in the RFP^+ cell was synchronized with adjacent cardiomyocytes, indicating electrical coupling. The plots correspond to scans within the coloured boxes shown in **h**. **j**, Caffeine augmentation of the amplitude of calcium oscillations in an RFP^+ cell. Blue staining, (**a–f**), DAPI. Scale bars, 50 μm (**a**) and 10 μm (**b–j**).

activation by tamoxifen concentrations compatible with maintenance of pregnancy. Co-staining for differentiation markers showed that β-gal^{+} cells differentiated into cardiomyocyte, endothelial and smooth muscle cell lineages (Fig. 3c–e). The distribution of labelled cells between these lineages was comparable between pulse and constitutive labelling approaches. We verified co-expression of cardiomyocyte markers and pulse-labelled lineage tracers at the single cell level by staining cardiomyocytes dissociated from E16.5 *Wt1*$^{CreERT2/+}$;*Z/Red* hearts, pulsed with tamoxifen at E10.5 and E11.5 (Fig. 3f–h). We consistently observed these pulse-labelled cardiomyocytes, although the frequency ($0.02 \pm 0.01\%$) was notably less than with constitutive labelling with *Wt1*^{GFPCre}. Within the window of the tamoxifen pulse, cardiac Wt1 expression was confined to the epicardium (Fig. 1). On the basis of these data, we conclude that a subset of epicardial cells expressing *Wt1* differentiate into cardiomyocytes.

To delimit the location of *Wt1*-expressing cells that differentiate into cardiomyocytes further, we microdissected E11.5 *Wt1*$^{GFPCre/+}$;*Z/Red* fetal hearts. Serial enzymatic digestion of intact hearts yielded epicardial cells preferentially in the early fractions due to their location on the exterior of the heart. Early (epicardial) and late (negative control) digestion fractions were sorted for GFP fluorescence, yielding a population enriched for active GFP expression (Fig. 3i). These GFP^{+} cells were plated onto either mitotically inactivated cardiac feeders (Fig. 3j) or untreated tissue culture dishes (Fig. 3k). In both conditions a subset of the sorted *Wt1*$^{GFPCre/+}$;*Z/Red* cells differentiated into cardiomyocytes, identified by co-expression of the RFP lineage tracer and cardiomyocyte markers (Fig. 3j–k). These data provide further evidence that a subset of heart cells actively expressing *Wt1*, confined within epicardium at E10–11.5 (Fig. 1), differentiate into cardiomyocytes.

We obtained further independent evidence that epicardial cells differentiate into cardiomyocytes by selective dye labelling of epicardium in E11.5 explanted hearts (Fig. 3l). Explanted hearts were briefly incubated in culture media containing the dye CMFDA and then placed in culture media without dye. This resulted in selective labelling of epicardium, as confirmed in sections of hearts fixed immediately after CMFDA incubation (Fig. 3l, 0 h). After two days of explant culture, labelled epicardial cells were found within the myocardium, and a subset expressed the cardiomyocyte marker Nkx2-5 (Fig. 3l, 48 h). The presence of dye and the cardiomyocyte markers Nkx2-5, Tnni3 and Actn1 within the same cell was further demonstrated in single cells isolated by dissociating heart explants two days after labelling (Fig. 3m). Collectively, these data indicate that precursors actively expressing *Wt1* within E10.5–E11.5 epicardium differentiate into cardiomyocytes.

Reported cardiac precursors are derived from multipotent *Isl1*$^{+}$ and *Nkx2-5*$^{+}$ progenitors[1–4]. We used Cre-based lineage tracing to determine whether *Wt1*$^{+}$ proepicardium cells are related to these progenitors or represent a different cardiogenic lineage. Using an *Nkx2-5*^{IRESCre} knock-in allele[18], we found that *Nkx2-5*-driven-Cre activated *Rosa26*^{fsLz} in a subset of proepicardium cells (Fig. 4a, b), suggesting that the descent of *Wt1*$^{+}$ proepicardium cells from *Nkx2-5*-expressing cells occurs.

We independently corroborated this result using a different *Nkx2-5* knock-in allele, *Nkx2-5*^{Cre} (ref. 19), and a Cre-activated reporter, *Gata4*^{flap}. Gata4 is expressed in cardiomyocyte, smooth muscle and

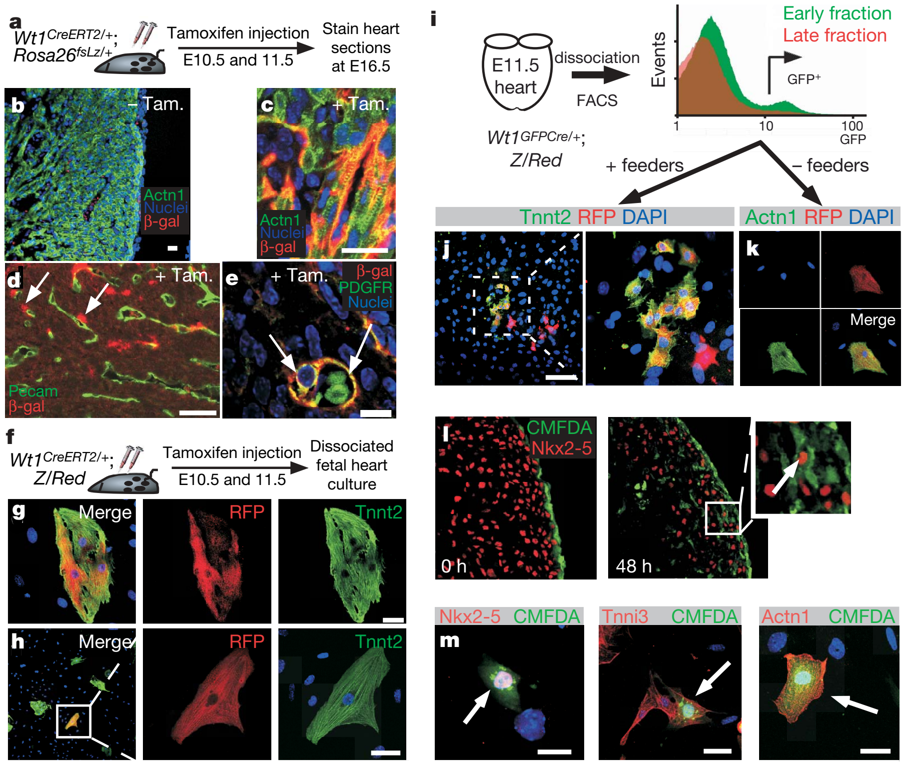


Figure 3 | *Wt1*-expressing epicardial cells differentiate into cardiomyocytes. a–h, Temporal restriction of Cre labelling with *Wt1*^{CreERT2}. *Wt1*^{CreERT2} did not recombine the fate mapping reporter *Rosa26*^{fsLz} in the absence of tamoxifen (Tam.) injection (**b**). Tamoxifen treatment at E10.5 and E11.5 induced *Wt1*^{CreERT2} labelling of cardiomyocytes, as determined in both tissue sections (**c**) and in dissociated heart culture (**g**–**h**). Also labelled were smooth muscle cells, marked by expression of the platelet derived growth factor receptor β (PDGFR), adjacent to endothelial tubes, marked by expression of platelet/endothelial cell adhesion molecule 1 (Pecam) (**d**, **e**) . **i**–**k**, E11.5 *Wt1*$^{GFPCre/+}$ heart cells actively expressing Wt1, as determined by GFP fluorescence, differentiated into cardiomyocytes. *Wt1*$^{+}$ epicardial cells were enriched in early digestion fractions compared to late digestion fractions (**i**, FACS analysis). GFP^{+} cells from early fractions were plated either with mitotically inactivated feeders (**j**) or without feeders (**k**). **l**, **m**, CMFDA dye, selectively incorporated into E11.5 epicardium, was found in cardiomyocytes at 48 h. Brief incubation of E11.5 heart explants with the dye CMFDA resulted in selective labelling of epicardium (0 h culture). After 48 h in explant culture, dye-labelled cells were present in the myocardial wall (**l**). A subset of dye-labelled cells co-expressed the cardiomyocyte markers Nkx2-5, cardiac troponin I (Tnni3) and Actn1 (**m**). Arrows indicate co-expression; scale bars represent 10 μm.

endothelial cell compartments of the myocardium, as well as in proepicardium (Supplementary Fig. 9)[20–22]. Therefore, within this domain, endogenous *Gata4* regulatory elements can be used to drive expression of a Cre-dependent reporter gene, alkaline phosphatase; we generated such a reporter, *Gata4^{flap}* (Supplementary Fig. 10). In the absence of Cre, *Gata4^{flap}* did not express alkaline phosphatase (Supplementary Fig. 10). In the presence of well-characterized Cre transgenes, *Gata4^{flap}* expressed alkaline phosphatase in patterns consistent with the expected sites of Cre activity (Cre driven by the *cardiac troponin* (*cTnTCre*) and *myosin heavy chain α* (*Myh6Cre*) promoters in myocardium, and by the *endothelial-specific receptor tyrosine kinase* promoter (*Tie2Cre*) in endothelium; Supplementary Fig. 10). Quantitative analysis revealed that *Myh6Cre* activated *Gata4^{flap}* in a greater percentage of cardiomyocytes than *Rosa26^{fsLz}* (*Gata4^{flap}* 93% ± 3%, versus *Rosa26^{fsLz}* 72% ± 6%, $P < 0.005$, $n = 4$), suggesting that *Gata4^{flap}* has greater sensitivity to Cre recombination than *Rosa26^{fsLz}*. Therefore, we addressed whether *Gata4^{flap}* would show a greater contribution of *Nkx2-5*$^+$ cells to proepicardium than suggested by *Nkx2-5^{IRESCre};Rosa26^{fsLz}*.

Gata4^{flap} demonstrated a robust contribution of *Nkx2-5*-expressing progenitors to proepicardium (Fig. 4c, asterisk). The *Nkx2-5^{Cre}*-labelled proepicardium cells expressed Wt1 (Fig. 4d), indicating that *Wt1*$^+$ cells in proepicardium are derived from *Nkx2-5*-expressing precursors. *Gata4^{flap}* also showed a robust contribution of *Isl1*-expressing precursors to the *Wt1*$^+$ cells in proepicardium (Supplementary Fig. 11b). Supporting this finding, Wt1 and Isl1 were expressed in adjacent regions at E8.0 and a subset of cells was positive for both markers (Supplementary Fig. 12a, b).

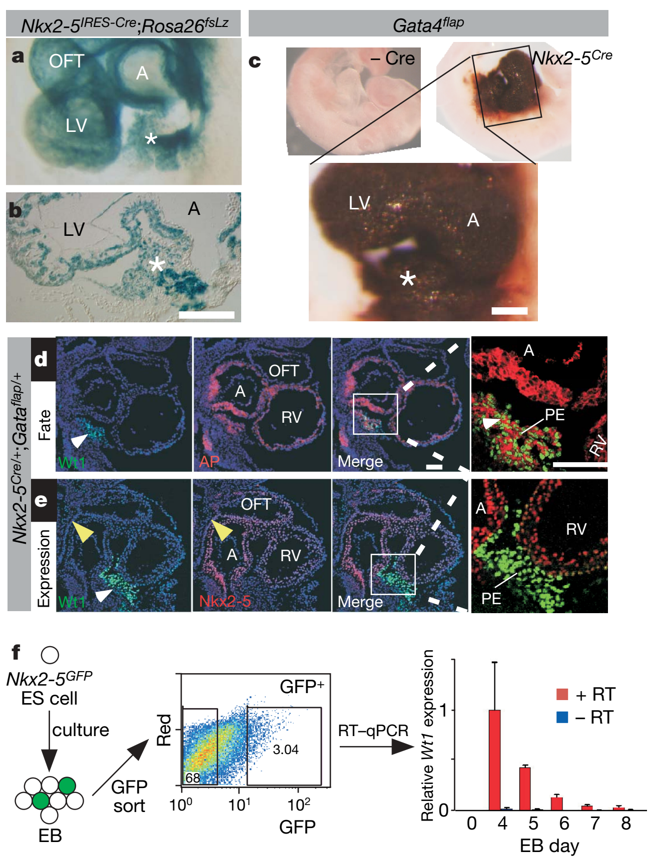


Figure 4 | Proepicardium arises from *Nkx2-5*$^+$ precursors. a–c, The *Nkx2-5* fate map was determined in *Nkx2-5^{IRESCre};Rosa26^{fsLz}* (**a**, **b**) or *Nkx2-5^{Cre};Gata4^{flap}* (**c**) embryos. *Nkx2-5*-driven Cre activated lineage tracer expression in proepicardium (asterisk). **d**, *Nkx2-5* fate, marked by the expression of alkaline phosphatase (AP, red membrane staining) compared to Wt1 immunohistochemistry (green nuclear staining). Wt1 was expressed in Nkx2-5-derived cells in proepicardium (white arrowhead). **e**, Immunohistochemistry comparing Nkx2-5 (red) and Wt1 expression (green). The yellow arrowhead indicates anterior heart field, which did not express Wt1. **f**, *Nkx2-5^{GFP}* embryonic stem (ES) cells were differentiated in embryoid body (EB) culture for the indicated number of days. The cells were then dissociated, and GFP$^+$ cells were isolated by flow cytometry. *Wt1* expression in GFP$^+$ cells was measured by quantitative reverse transcription PCR (RT–qPCR). Bars represent mean ± s.d; $n = 3$. Scale bars, 50 μm. A, atrium; RV, right ventricle; LV, left ventricle; OFT, outflow tract.

Although *Wt1*$^+$ cells in proepicardium were labelled by *Nkx2-5*-driven Cre, they did not actively co-express Nkx2-5 at E9.5 (Fig. 4e). In E8.0 embryos, Nkx2-5 and Wt1 were expressed in adjacent cells, but were not co-expressed (Supplementary Fig. 12c), suggesting that *Nkx2-5* and *Wt1* are expressed sequentially, or transiently co-expressed. To investigate the relationship between *Nkx2-5* and *Wt1* expression further, we studied the expression of *Wt1* in *Nkx2-5*$^+$ cells during embryoid body differentiation of embryonic stem cells. Using transgenic *Nkx2-5^{GFP}* embryonic stem cells[4] and fluorescence-activated cell sorting (FACS), we isolated Nkx2-5-expressing cells at several time points during embryoid body differentiation. *Wt1* was transiently upregulated in *Nkx2-5*$^+$ cells during embryonic stem cell differentiation (Fig. 4f). This result was specific, as we did not detect significant *Wt1* expression in parallel experiments with *Mef2c-AHF-GFP* embryonic stem cells[23] (data not shown). In embryos, the activity domain of the *Mef2c-AHF* enhancer (which was 'on' in anterior heart field, 'off' in proepicardium) did not overlap with the Wt1 expression domain[24] (Fig. 4e, yellow arrowhead). Collectively, these data suggest that *Nkx2-5* and *Wt1* are either sequentially expressed or transiently co-expressed in a subset of proepicardium precursors.

We have shown that *Wt1*$^+$ proepicardium and epicardial cells contribute to the cardiomyocyte lineage during normal heart development (Supplementary Fig. 1). *Wt1*$^+$ cells located on the heart at E10.5–E11.5 differentiate into functional cardiomyocytes. Although differentiation of proepicardium cells into cardiomyocytes was previously noted *in vitro*[25], previous fate-mapping studies of proepicardium cells, using retroviral labels in chick or transgene labels in mice, did not describe proepicardium contribution to the cardiomyocyte lineage *in vivo*[8–10,12,13]. This may be attributable to differences in methodology, species or domains of transgene activity. Consistent with the capacity of *Wt1*-expressing cells to differentiate into cardiomyocytes, *Wt1*$^+$ proepicardium and epicardial cells are derived from progenitors that express *Nkx2-5* and *Isl1*, suggesting that they share a common developmental origin with previously described multipotent cardiogenic progenitors[1,2]. These experiments have identified a previously unrecognized cardiomyocyte progenitor population in the developing heart that may be of use for cardiac regeneration or repair.

METHODS SUMMARY

Gene targeting and mouse lines are described in Methods. Fetal hearts were dissociated by serial digestion with collagenase and trypsin. For calcium imaging, dissociated fetal heart cultures were loaded with Fluo-4 AM and imaged with an Olympus FV1000 confocal microscope. Cardiac feeders were prepared and mitotically inactivated as described[26]. *Nkx2-5^{GFP}* and *Mef2c-AHF-GFP* embryonic stem cells were differentiated in embryoid body culture and sorted for GFP fluorescence as described previously[4,23]. Immunohistochemistry was performed according to standard methods, using primary and secondary antibodies listed in Supplementary Table 1.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 March; accepted 8 May 2008.
Published online 22 June 2008.

1. Martin-Puig, S., Wang, Z. & Chien, K. R. Lives of a heart cell: tracing the origins of cardiac progenitors. *Cell Stem Cell* **2**, 320–331 (2008).
2. Laugwitz, K. L., Moretti, A., Caron, L., Nakano, A. & Chien, K. R. Islet1 cardiovascular progenitors: a single source for heart lineages? *Development* **135**, 193–205 (2008).
3. Moretti, A. *et al.* Multipotent embryonic *Isl1*$^+$ progenitor cells lead to cardiac, smooth muscle, and endothelial cell diversification. *Cell* **127**, 1151–1165 (2006).

4. Wu, S. M. *et al.* Developmental origin of a bipotential myocardial and smooth muscle cell precursor in the mammalian heart. *Cell* **127,** 1137–1150 (2006).
5. Manner, J., Perez-Pomares, J. M., Macias, D. & Munoz-Chapuli, R. The origin, formation and developmental significance of the epicardium: a review. *Cells Tissues Organs* **169,** 89–103 (2001).
6. Wessels, A. & Perez-Pomares, J. M. The epicardium and epicardially derived cells (EPDCs) as cardiac stem cells. *Anat. Rec. A* **276,** 43–57 (2004).
7. Smart, N. *et al.* Thymosin β4 induces adult epicardial progenitor mobilization and neovascularization. *Nature* **445,** 177–182 (2007).
8. Wilm, B., Ipenberg, A., Hastie, N. D., Burch, J. B. & Bader, D. M. The serosal mesothelium is a major source of smooth muscle cells of the gut vasculature. *Development* **132,** 5317–5328 (2005).
9. Merki, E. *et al.* Epicardial retinoid X receptor α is required for myocardial growth and coronary artery formation. *Proc. Natl Acad. Sci. USA* **102,** 18455–18460 (2005).
10. Gittenberger-de Groot, A. C., Vrancken Peeters, M. P., Mentink, M. M., Gourdie, R. G. & Poelmann, R. E. Epicardium-derived cells contribute a novel population to the myocardial wall and the atrioventricular cushions. *Circ. Res.* **82,** 1043–1052 (1998).
11. Manner, J. Does the subepicardial mesenchyme contribute myocardioblasts to the myocardium of the chick embryo heart? A quail–chick chimera study tracing the fate of the epicardial primordium. *Anat. Rec.* **255,** 212–226 (1999).
12. Dettman, R. W., Denetclaw, W. J., Ordahl, C. P. & Bristow, J. Common epicardial origin of coronary vascular smooth muscle, perivascular fibroblasts, and intermyocardial fibroblasts in the avian heart. *Dev. Biol.* **193,** 169–181 (1998).
13. Mikawa, T. & Gourdie, R. G. Pericardial mesoderm generates a population of coronary smooth muscle cells migrating into the heart along with ingrowth of the epicardial organ. *Dev. Biol.* **174,** 221–232 (1996).
14. Le, Y., Miller, J. L. & Sauer, B. GFP*cre* fusion vectors with enhanced expression. *Anal. Biochem.* **270,** 334–336 (1999).
15. Mao, X., Fujiwara, Y. & Orkin, S. H. Improved reporter strain for monitoring Cre recombinase-mediated DNA excisions in mice. *Proc. Natl Acad. Sci. USA* **96,** 5037–5042 (1999).
16. Vintersten, K. *et al.* Mouse in red: red fluorescent protein expression in mouse ES cells, embryos, and adult animals. *Genesis* **40,** 241–246 (2004).
17. Feil, R., Wagner, J., Metzger, D. & Chambon, P. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem. Biophys. Res. Commun.* **237,** 752–757 (1997).
18. Stanley, E. G. *et al.* Efficient Cre-mediated deletion in cardiac progenitor cells conferred by a 3′UTR-ires-Cre allele of the homeobox gene Nkx2–5. *Int. J. Dev. Biol.* **46,** 431–439 (2002).
19. Moses, K. A., DeMayo, F., Braun, R. M., Reecy, J. L. & Schwartz, R. J. Embryonic expression of an *Nkx2–5/Cre* gene using *ROSA26* reporter mice. *Genesis* **31,** 176–180 (2001).
20. Rivera-Feliciano, J. *et al.* Development of heart valves requires *Gata4* expression in endothelial-derived cells. *Development* **133,** 3607–3618 (2006).
21. Heikinheimo, M., Scandrett, J. M. & Wilson, D. B. Localization of transcription factor GATA-4 to regions of the mouse embryo involved in cardiac development. *Dev. Biol.* **164,** 361–373 (1994).
22. Watt, A. J., Battle, M. A., Li, J. & Duncan, S. A. GATA4 is essential for formation of the proepicardium and regulates cardiogenesis. *Proc. Natl Acad. Sci. USA* **101,** 12573–12578 (2004).
23. Qyang, Y. *et al.* The renewal and differentiation of *Isl1*$^{+}$ cardiovascular progenitors are controlled by a Wnt/β-Catenin Pathway. *Cell Stem Cell* **1,** 165–179 (2007).
24. Verzi, M. P., McCulley, D. J., De Val, S., Dodou, E. & Black, B. L. The right ventricle, outflow tract, and ventricular septum comprise a restricted expression domain within the secondary/anterior heart field. *Dev. Biol.* **287,** 134–145 (2005).
25. Kruithof, B. P. *et al.* BMP and FGF regulate the differentiation of multipotential pericardial mesoderm into the myocardial or epicardial lineage. *Dev. Biol.* **295,** 507–522 (2006).
26. Laugwitz, K. L. *et al.* Postnatal isl1+ cardioblasts enter fully differentiated cardiomyocyte lineages. *Nature* **433,** 647–653 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was funded by the National Heart, Lung and Blood Institute of the National Institutes of Health, USA, the American Heart Association, and by a charitable donation from E. P. Marram and K. K. Carpenter. We thank the Schwartz, Harvey, Schneider, Yanagisawa, Evans, Soriano, Orkin and Nagy laboratories for contributing mouse strains used in this study.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to W.T.P. (wpu@enders.tch.harvard.edu).

METHODS

Gene targeting. *Wt1*^{GFPCre}, *Wt1*^{CreERT2} and *Gata4*^{flap} targeting vectors were constructed using Red/ET recombineering[27]. The *Wt1* targeting strategy removed the first coding exon. The *Gata4*^{flap} targeting strategy replaced the first coding exon with a loxP-*Gata4* cDNA-transcription stop-loxP-alkaline phosphatase cDNA-Frt-pgkneo-Frt cassette. Gene targeting was confirmed by Southern blotting and PCR. Chimaeric mice were mated to *ActB::Flpe* mice[28] to remove the neomycin resistance cassette.

Mice. Nkx2-5^{Cre}, Nkx2-5^{IRESCre}, Isl1^{Cre}, Myh6αCre, cTntCre, Tie2Cre, Z/Red and Rosa26^{fsLz} mice were described previously[15,16,18,19,29–31]. *Wt1*^{CreERT2} was induced with tamoxifen. Tamoxifen (Sigma) was dissolved in ethanol at 100 mg ml^{-1} and then emulsified in sesame oil to a final concentration of 12.5 mg ml^{-1}. Two milligrams were injected intraperitoneally at E10.5 and E11.5. Mice were used in accordance with protocols approved by the Institutional Animal Care and Use Committee.

Tissue culture. Heart dissociation was performed by serial digestion with collagenase and trypsin, as described for neonatal hearts[32]. Cardiomyocytes were cultured on fibronectin-treated tissue culture dishes or glass coverslips using media M199 containing 10% fetal bovine serum, penicillin and streptomycin. For explant culture, hearts were excised and incubated in 1 μM CMFDA (Invitrogen) for 15 min at 37 °C. Hearts were washed twice and then incubated in DMEM containing 10% FBS, penicillin and streptomycin.

RNA expression. RNA expression was measured by quantitative real-time PCR using Sybr Green. Primers for *Wt1* were: forward, GCCTTCACCT TGCACTTCTC; reverse, GACCGTGCTGTATCCTTGGT. Gene expression was normalized to GAPDH, detected using proprietary primers and probes (Applied Biosystems). *Wt1 in situ* hybridization with digoxigenin-labelled probes was performed as described previously[33].

Immunohistochemistry. All procedures were performed according to standard protocols from Vector Laboratories and Jackson Immunoresearch or as described previously[34]. The sources and dilutions of primary and secondary antibodies are listed in Supplementary Table 1. The Cx43 antibody was a gift from D. Gutstein[35]. Weak signals were detected using a secondary antibody conjugated to polymerized peroxidase (Invitrogen) and then incubation with substrate DAB (Vector), tyramide-Cy3 (Perkin-Elmer) or tyramide-Alexa488 (Invitrogen).

β-galactosidase and alkaline phosphatase activity were detected as described previously[36]. Alkaline phosphatase was detected with either BCIP/NBT or Permanent Red (Dako). Permanent Red fluorescence was detected in the Cy5 channel of an FV1000 confocal microscope. Immunohistochemical studies of *Wt1*^{CreERT2} labelling were performed blinded to tamoxifen treatment group.

27. Liu, P., Jenkins, N. A. & Copeland, N. G. A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13,** 476–484 (2003).
28. Rodriguez, C. I. *et al.* High-efficiency deleter mice show that FLPe is an alternative to Cre-*loxP*. *Nature Genet.* **25,** 139–140 (2000).
29. Yang, L. *et al.* Isl1Cre reveals a common Bmp pathway in heart and limb development. *Development* **133,** 1575–1585 (2006).
30. Soriano, P. Generalized *lacZ* expression with the ROSA26 Cre reporter strain. *Nature Genet.* **21,** 70–71 (1999).
31. Jiao, K. *et al.* An essential role of Bmp4 in the atrioventricular septation of the mouse heart. *Genes Dev.* **17,** 2362–2367 (2003).
32. Pu, W. T., Ma, Q. & Izumo, S. NFAT transcription factors are critical survival factors that inhibit cardiomyocyte apoptosis during phenylephrine stimulation *in vitro*. *Circ. Res.* **92,** 725–731 (2003).
33. Brent, A. E., Schweitzer, R. & Tabin, C. J. A somitic compartment of tendon progenitors. *Cell* **113,** 235–248 (2003).
34. Zhou, B. *et al.* G-CSF-mobilized peripheral blood mononuclear cells from diabetic patients augment neovascularization in ischemic limbs but with impaired capability. *J. Thromb. Haemost.* **4,** 993–1002 (2006).
35. Gutstein, D. E., Liu, F. Y., Meyers, M. B., Choo, A. & Fishman, G. I. The organization of adherens junctions and desmosomes at the cardiac intercalated disc is independent of gap junctions. *J. Cell Sci.* **116,** 875–885 (2003).
36. Lobe, C. G. *et al.* Z/AP, a double reporter for Cre-mediated recombination. *Dev. Biol.* **208,** 281–292 (1999).

LETTERS

Functional asymmetry in *Caenorhabditis elegans* taste neurons and its computational role in chemotaxis

Hiroshi Suzuki[1]*†, Tod R. Thiele[2]*, Serge Faumont[2], Marina Ezcurra[3], Shawn R. Lockery[2]* & William R. Schafer[1,3]†

Chemotaxis in *Caenorhabditis elegans*, like chemotaxis in bacteria[1], involves a random walk biased by the time derivative of attractant concentration[2,3], but how the derivative is computed is unknown. Laser ablations have shown that the strongest deficits in chemotaxis to salts are obtained when the ASE chemosensory neurons (ASEL and ASER) are ablated, indicating that this pair has a dominant role[4]. Although these neurons are left–right homologues anatomically, they exhibit marked asymmetries in gene expression and ion preference[5–7]. Here, using optical recordings of calcium concentration in ASE neurons in intact animals, we demonstrate an additional asymmetry: ASEL is an ON-cell, stimulated by increases in NaCl concentration, whereas ASER is an OFF-cell, stimulated by decreases in NaCl concentration. Both responses are reliable yet transient, indicating that ASE neurons report changes in concentration rather than absolute levels. Recordings from synaptic and sensory transduction mutants show that the ON–OFF asymmetry is the result of intrinsic differences between ASE neurons. Unilateral activation experiments indicate that the asymmetry extends to the level of behavioural output: ASEL lengthens bouts of forward locomotion (runs) whereas ASER promotes direction changes (turns). Notably, the input and output asymmetries of ASE neurons are precisely those of a simple yet novel neuronal motif for computing the time derivative of chemosensory information, which is the fundamental computation of *C. elegans* chemotaxis[3,8]. Evidence for ON and OFF cells in other chemosensory networks[9–12] suggests that this motif may be common in animals that navigate by taste and smell.

To image the activity of ASE neurons in response to stepwise changes in NaCl concentration, we used the genetically encoded calcium sensor cameleon[13], which reports increases in calcium concentration as increases in the ratio of fluorescence emission at distinct wavelengths. We observed that a stepwise increase (up-step) in NaCl concentration evoked a rapid increase in emission ratio in ASE left (ASEL) (Fig. 1a). Similar results were obtained when up-steps were delivered from different baseline NaCl concentrations (Supplementary Fig. 1); emission ratio increases were absent when a calcium-insensitive form of the cameleon was used (Supplementary Fig. 2). Thus, ASEL seemed to be activated by an increase in NaCl concentration.

ASE right (ASER) neurons showed the opposite pattern of response. A stepwise decrease (down-step) in NaCl concentration evoked a rapid increase in emission ratio, and similar results were obtained when down-steps were delivered from different baselines (Fig. 1b, Supplementary Fig. 1). In addition, we observed that an up-step of NaCl evoked a prominent decrease in the emission ratio in ASER (Fig. 1a, Supplementary Fig. 1). Neither type of ASER response was evident when we used calcium-insensitive cameleon (Supplementary Fig. 2). Thus, ASER seemed to be activated by a decrease in NaCl concentration and de-activated by an increase in NaCl concentration.

Opposing responses in ASEL and ASER were observed consistently across a range of step durations (from 10 s to 60 s, data not shown) and for a variety of salts, including those that did or did not contain Na^+ and Cl^- ions (Supplementary Fig. 2). In dose-response experiments, we found that the magnitude of the responses in ASEL and ASER was a saturating function of step amplitude (Fig. 1c, d), and that ASER is more sensitive than ASEL to small changes in NaCl (Fig. 1c, d). We conclude that opposing responses are a general feature of ASE neurons.

Together, these data indicate that the ASEL and ASER neurons function like ON-cells and OFF-cells, respectively—a common coding mechanism in visual systems[14]. Notably, the ON and OFF responses to preferred stimuli were transient in the face of a maintained concentration change. These responses therefore signal changes in salt concentration rather than its absolute level and thus provide the basis for computing the time derivative of concentration.

ASE neurons are reported to have different ion sensitivities, in that ablating ASEL impairs chemotaxis mainly to Na^+ ions whereas ablating ASER impairs chemotaxis mainly to Cl^- ions[6]. To assess ion sensitivity directly, we imaged the response of ASEL and ASER to 10-mM step changes in sodium acetate and ammonium chloride (Fig. 1e and Supplementary Fig. 3); at this step size, neither cell responded to ammonium or acetate ions (see also Supplementary Fig. 2b). We found that only ASEL responded to Na^+ ions, a result that is consistent with the effects on Na^+ chemotaxis when ASEL or ASER is ablated. In addition, we found that ASER responded much more strongly to Cl^- ions than did ASEL. This result is consistent with the strong effect on Cl^- chemotaxis when ASER is ablated. The comparatively weak response of ASEL to chloride may or may not be consistent with the absence of an effect on Cl^- chemotaxis when ASEL is ablated; this would depend on the relative sensitivity of imaging and behavioural experiments, which is not known. Overall, we conclude that ASE neurons are differentially sensitive to Na^+ and Cl^- ions.

In principle, the opposing responses of ASEL and ASER neurons could result from intrinsic differences between the sensory properties of ASEL and ASER or from intrinsic differences in chemosensory neurons that are presynaptic to ASE neurons. We therefore imaged ASEL and ASER in *unc-13*(*e51*) and *snb-1*(*md247*) mutants, which have impairments in the release of synaptic vesicles[15,16]. In addition, we imaged ASE in *unc-31*(*e928*) mutants, which are defective for dense-core vesicle release[17]. We observed that the ON–OFF asymmetry of ASEL and ASER was preserved in each of these mutants (Fig. 2a, b and Supplementary Fig. 4), suggesting that the functional asymmetry of ASE neurons is probably intrinsic rather than synaptic in origin.

[1]Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093, USA. [2]Institute of Neuroscience, University of Oregon, Eugene, Oregon 97403, USA. [3]MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK. †Present addresses: Center for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario M5S 3H2, Canada (H.S.); MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, UK (W.R.S.).
*These authors contributed equally to this work.

The mechanism of sensory transduction in ASE neurons is unknown, but the cGMP-dependent pathway is a strong candidate in two respects. First, salt chemotaxis is impaired by mutations in the genes *tax-2* and *tax-4*, which are expressed in ASE neurons and encode subunits of a cGMP-gated cation channel[18,19]. Second, salt chemotaxis is also impaired by mutations in the gene *egl-4*, which is expressed in ASE neurons and encodes a cGMP-dependent protein kinase (PKG)[20,21]. When we imaged calcium responses to NaCl up-steps and down-steps in ASEL and ASER in both *tax-2*(*p671*) and *tax-4*(*p678*) mutants, we found that the absence of either *tax-2* or *tax-4* function eliminated all responses in ASE neurons (Fig. 2c and Supplementary Fig. 4). These results support a model in which ASE responses are mediated by cGMP signalling. The ON–OFF asymmetry of ASE neurons could be explained if cGMP levels are increased by NaCl up-steps in ASEL and by down-steps in ASER. ASE responses were also completely absent in *egl-4*(*n479*) mutants (Fig. 2d), indicating that PKG is also required for salt detection. EGL-4 regulates olfactory adaptation involving phosphorylation of the TAX-2 protein in other *C. elegans* neurons[22]. Thus, salt adaptation and detection may be linked in ASE neurons.

Behavioural studies have shown that NaCl up-steps increase the probability of forward locomotion and, concomitantly, decrease the probability of turning, whereas NaCl down-steps have the opposite effects[8]. The fact that ASEL and ASER are strongly activated by up-steps and down-steps, respectively (Fig. 1a, b), suggests that ASEL contributes to up-step behaviour whereas ASER contributes to down-step behaviour. To test this model, we selectively activated

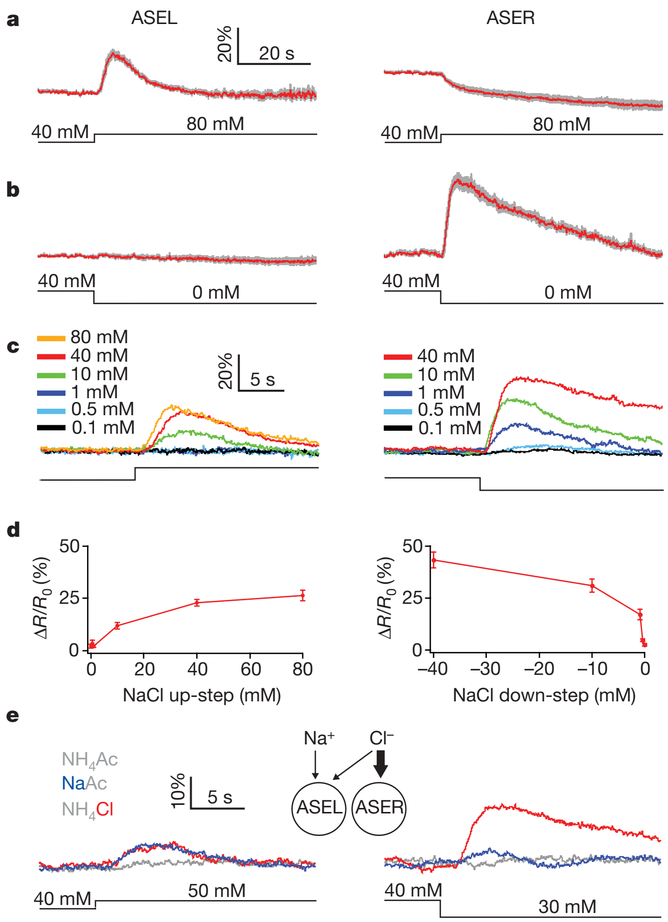


Figure 1 | Response of ASEL and ASER to NaCl concentration steps. a, **b**, Average calcium transients in ASE neurons in response to NaCl concentration steps of ±40 mM from a baseline of 40 mM. The grey band represents ±1 standard error of the mean (s.e.m.). Transients were imaged using the ratiometric calcium sensor cameleon, and traces indicate average percentage change in R/R_0, where R is the fluorescence emission ratio at a given time point and R_0 is its initial value. ASEL is shown on the left and ASER is shown on the right. The concentration step is indicated below the calcium traces; $n \geq 5$ recordings, with one recording per worm. **c**, Average calcium transients in response to concentration steps of various amplitudes from a baseline of 40 mM. The trace colour denotes step size (in mM); s.e.m. is not shown. **d**, Summary of the data in **c**, showing the effect of step size on average peak response amplitude; error bars represent ±s.e.m. **e**, Differential sensitivity of ASEL and ASER to Na^+ and Cl^-. Traces are the average percentage change in calcium signal. Trace colour denotes stimulus condition as shown in the figure; $n \geq 5$ recordings, with one recording per worm. See Supplementary Fig. 3 for s.e.m. The diagram depicts the relative ion sensitivities inferred from the imaging data.

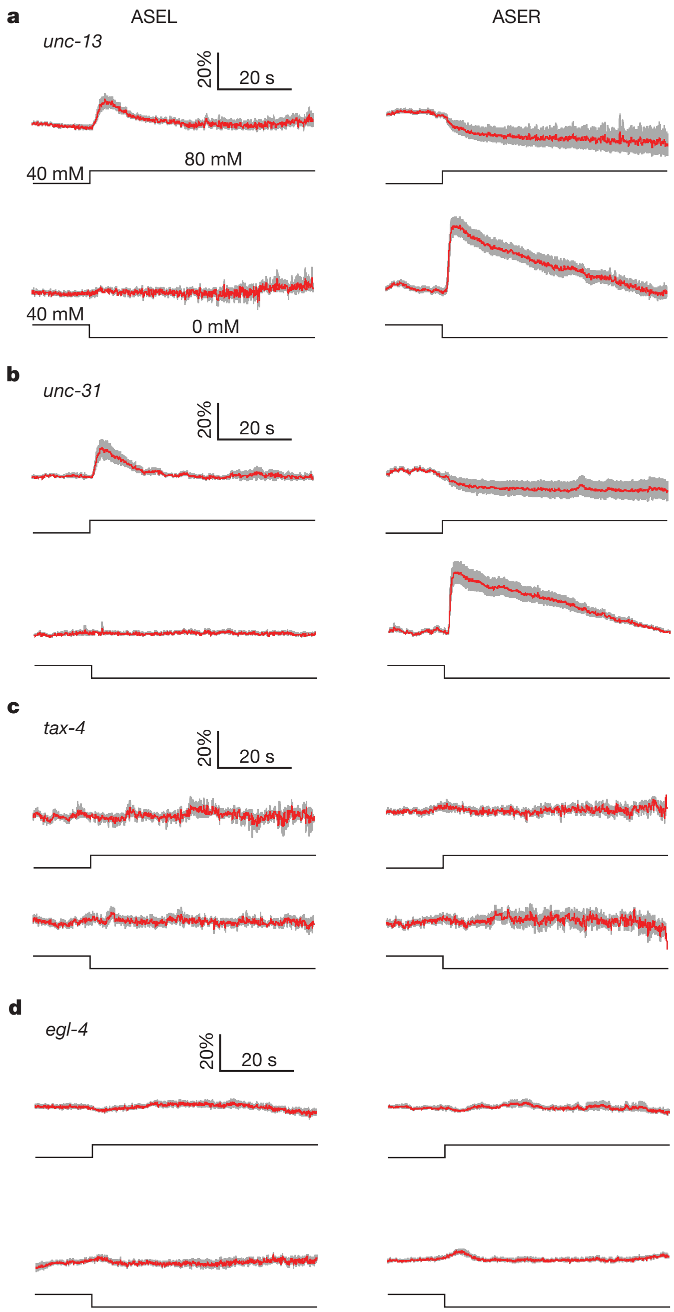


Figure 2 | Effects of synaptic and signal transduction mutants on ASE sensory responses. Average ASE calcium transients in four mutant strains in response to NaCl concentration steps of ±40 mM from a baseline of 40 mM. **a**, **b**, Mutants with defects in synaptic vesicle release. **c**, **d**, Mutants with defects in cGMP-dependent signalling. In each panel, traces indicate the average percentage change in ratio; the grey band represents ±1 s.e.m. ASEL is shown on the left and ASER is shown on the right, the concentration step for each trace is as in **a**. $n \geq 5$ recordings, with one recording per worm.

either ASEL or ASER by means of transgenic strains in which the mammalian cation channel TRPV1, which opens in response to the exogenous ligand capsaicin, is expressed either in ASEL or in ASER[23]. Calcium imaging showed that only the ASE neuron expressing TRPV1 was activated by capsaicin (Fig. 3a, b and Supplementary Fig. 5). We found that activation of ASEL increased forward probability (Fig. 3c; analysis of variance (ANOVA), $P < 0.01$) whereas activation of ASER had the opposite effect (Fig. 3d; ANOVA, $P < 0.05$). We conclude that ASEL activation causes runs whereas ASER activation causes turns; thus, the functional asymmetry between ASE neurons extends to the level of behavioural output.

The question remained, however, whether ASE neurons make a necessary contribution to the run and turn behaviours that underlie chemotaxis to salts such as NaCl[3,8]. When ASER was ablated and animals were tested with its preferred stimulus (a down-step), we observed a large deficit in turn behaviour (Fig. 4c; ANOVA, $P < 10^{-3}$; see also Supplementary Fig. 6); the presence of residual turn behaviour is consistent with the observation that chemosensory neurons other than ASE neurons contribute to chemotaxis[4]. Conversely, when ASER was ablated and animals were tested with its non-preferred stimulus (an up-step), we observed a small but significant deficit in run behaviour (Fig. 4b; ANOVA, $P < 0.01$). In confirmation, we found that when ASER was ablated together with ASEL, the defect in the response to up-steps was greater than when ASEL alone was ablated (Fig. 4f versus 4d; ANOVA, $P < 0.05$). We conclude that ASER makes a necessary contribution to runs and turns. Notably, because ASER is deactivated by its non-preferred stimulus (Figs 1a and 4a), the positive contribution of ASER to run behaviour in intact animals must be the result of de-suppression of forward probability. This finding suggests that ASER is tonically active at the baseline NaCl concentration, like at least one olfactory neuron in *C. elegans*[24].

When ASEL was ablated and animals were tested with its preferred stimulus (an up-step), we observed a significant deficit in run behaviour (Fig. 4d; ANOVA, $P < 10^{-3}$; see also Supplementary Fig. 6); the presence of residual run behaviour is, again, consistent with the observation that other chemosensory neurons contribute to chemotaxis[4]. In contrast, when ASEL was ablated and animals were tested with its non-preferred stimulus (a down-step), we found no deficit in turn behaviour (Fig. 4e; ANOVA, $P = 0.13$); also, ablating ASEL together with ASER did not enhance the deficit produced by ablating ASER alone (Fig. 4c versus 4g; ANOVA, $P = 0.50$). Thus, ASEL makes a necessary contribution to runs but not to turns.

We have shown that a left–right homologous pair of chemosensory neurons is functionally asymmetric at the cellular level: ASEL acts as an ON-cell, is specialized for Na^+ sensation and causes runs, whereas

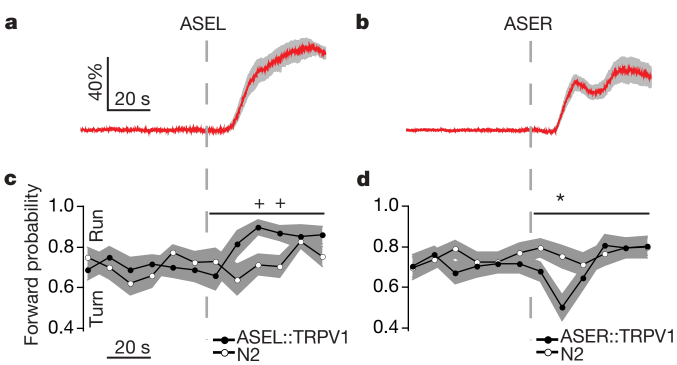


Figure 3 | Unilateral activation of ASEL and ASER. a, b, Average capsaicin-induced calcium transients in strains expressing the capsaicin receptor TRPV1 exclusively in either ASEL or ASER. The dashed line indicates the onset of capsaicin application, which was maintained for the duration of the experiment (ASEL, 25 μM, $n = 6$; ASER, 5 μM, $n = 7$). The grey band represents ±1 s.e.m. **c, d,** Behavioural effects of capsaicin in the strains shown in **a** and **b** at the same capsaicin concentrations. The probability of forward locomotion is plotted against time. An increase in forward probability is termed a run whereas a decrease in forward probability is termed a turn. Statistical significance (TRPV1 strain versus wild-type N2) was assessed by means of a repeated measures ANOVA over a 1-min window (horizontal line above traces; *P* values are given in the text) after capsaicin application. Asterisk, significant difference between the means at each time point after correcting for multiple comparisons (*t*-test, $P < 0.05$); plus symbols, significant differences detected in uncorrected *t*-tests ($P < 0.05$). Imaging and behavioural data are from different individuals. The grey band represents ±1 s.e.m. with $n \geq 22$.

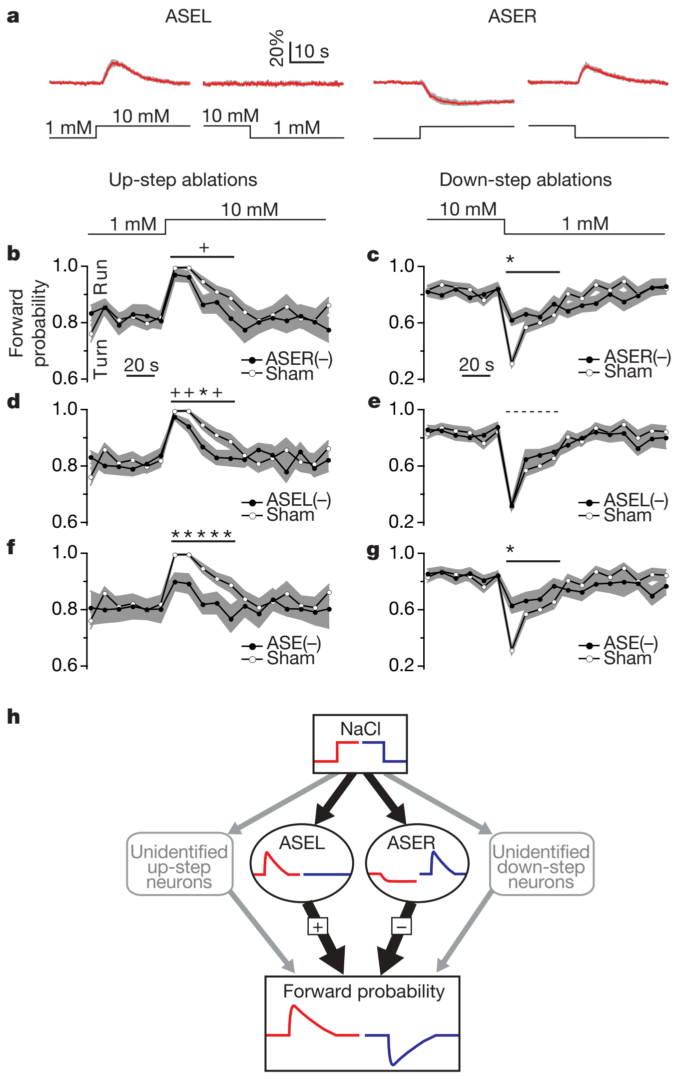


Figure 4 | Roles of ASEL and ASER in NaCl step-response behaviour. a, Average ASE calcium transients in response to concentration steps of ±9 mM. The concentration step is indicated below the first calcium trace. The grey band represents ±1 s.e.m.; $n \geq 5$ recordings, with one recording per worm. **b–g,** Effects of unilateral and bilateral ASE ablations on the behavioural response to concentration steps. The probability of forward locomotion is plotted against the time relative to the step. Statistical significance (ablation versus sham operated) was assessed by means of a repeated measures ANOVA over the indicated time window (horizontal lines above the traces) after the step (shown above **b** and **c**). ASEL(–), ASEL ablation; ASER(–), ASER ablation; ASE(–), bilateral ASE ablation; solid horizontal line, ANOVA significant at $P < 0.05$ or less; dashed horizontal line, not significant; asterisks, time points at which there was a significant difference between means after correcting for multiple comparisons (*t*-test; $P < 0.05$); plus symbols, time points at which there were significant differences in uncorrected *t*-tests ($P < 0.05$). Imaging and behavioural data are from different individuals. The grey band represents ±1 s.e.m., with $n \geq 15$ in each panel. **h,** Functional connectivity implied by **b–g** together with imaging data (panel **a** and Fig. 1a, b) and unilateral activation experiments (Fig. 3). Unidentified neurons (shown in grey) account for residual behaviour when ASE is ablated.

ASER acts as an OFF-cell, is specialized for Cl^- detection and causes turns. ON-cells and OFF-cells are a common feature of early stages in visual processing[14]. ON-cell chemosensory neurons are also well known, as are chemosensory neurons that hyperpolarize in response to concentration increases[25,26] and chemosensory OFF-cells, including the AWC olfactory neurons in *C. elegans*[9–12,24]. However, the observation of a single pair of anatomically homologous chemosensory neurons with ON–OFF functionality and differential chemical sensitivity seems to be unprecedented.

The asymmetry of ASE neurons at the sensory level (ON-cell versus OFF-cell) and the behavioural level (runs versus turns) immediately suggests that the time derivative is computed, at least in part, by a two-stage neuronal motif (Fig. 4h). First, each ASE neuron is transiently activated by its preferred stimulus, such that the time course of ASEL activation approximates the derivative of an up-step whereas the time course of ASER activation approximates the inverse derivative of a down-step. Second, the effects of ASE activation converge, but with opposing effects: ASEL positively regulates forward locomotion whereas ASER negatively regulates it, causing turns. Thus, at the point of convergence, the net effect of ASE activation is a behavioural signal that approximates the time derivative of salt concentration. The functional asymmetries of ASE neurons may have emerged as a means of computing a quantity that is essential to chemotaxis in this organism. ASE asymmetry is established and maintained by a complex gene regulatory network[7]; the critical role of chemotaxis in the search for food and habitat[27,28] would seem to justify the complexity of this network.

METHODS SUMMARY

Calcium imaging. Animals expressing cameleon YC2.12 in ASEL and ASER neurons were glued to coverslips and submerged in a pool of saline. NaCl steps were produced by placing the worm in a plume between an inflow and an outflow pipette; the solution feeding the plume was changed by valves controlled by the data acquisition system. For ratiometry, images in two distinct wavelength bands were projected in juxtaposition on the photodetector by means of a beam splitter.

Selective activation of ASE neurons. The capsaicin receptor TRPV1 was expressed exclusively in ASEL or ASER using the *gcy-7* and *gcy-5* promoters, respectively[5]. Worms were glued to coverslips, but here the tail was free to move. Capsaicin was delivered as above. Behaviour was scored manually by pressing computer keys to record bouts of forward swimming[29]; the observer was blind to the genotype.

Neuronal ablations. ASE neurons were identified by position and were ablated with a focused laser beam[30]. Behaviour was tested using the concentration-clamp assay[8] in which a worm is placed on a porous membrane that is infused from below by a pair of inverted showerheads that emit osmotically balanced solutions with different NaCl concentrations; concentration changes are produced by sliding the showerhead assembly relative to the worm. Behaviour was scored as above.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 January; accepted 20 March 2008.

1. Segall, J. E., Block, S. M. & Berg, H. C. Temporal comparisons in bacterial chemotaxis. *Proc. Natl Acad. Sci. USA* **83,** 8987–8991 (1986).
2. Dusenbery, D. B. Responses of the nematode *Caenorhabditis elegans* to controlled chemical stimulation. *J. Comp. Physiol.* **136,** 127–331 (1980).
3. Pierce-Shimomura, J. T., Morse, T. M. & Lockery, S. R. The fundamental role of pirouettes in *Caenorhabditis elegans* chemotaxis. *J. Neurosci.* **19,** 9557–9569 (1999).
4. Bargmann, C. I. & Horvitz, H. R. Chemosensory neurons with overlapping functions direct chemotaxis to multiple chemicals in *C. elegans. Neuron* **7,** 729–742 (1991).
5. Yu, S., Avery, L., Baude, E. & Garbers, D. L. Guanylyl cyclase expression in specific sensory neurons: a new family of chemosensory receptors. *Proc. Natl Acad. Sci. USA* **94,** 3384–3387 (1997).
6. Pierce-Shimomura, J. T., Faumont, S., Gaston, M. R., Pearson, B. J. & Lockery, S. R. The homeobox gene *lim-6* is required for distinct chemosensory representations in *C. elegans. Nature* **410,** 694–698 (2001).
7. Johnston, R. J. Jr, Chang, S., Etchberger, J. F., Ortiz, C. O. & Hobert, O. MicroRNAs acting in a double-negative feedback loop to control a neuronal cell fate decision. *Proc. Natl Acad. Sci. USA* **102,** 12449–12454 (2005).
8. Miller, A. C., Thiele, T. R., Faumont, S., Moravec, M. L. & Lockery, S. R. Step-response analysis of chemotaxis in *Caenorhabditis elegans. J. Neurosci.* **25,** 3369–3378 (2005).
9. Takagi, S. F. & Shibuya, T. 'On'- and 'Off'-responses of the olfactory epithelium. *Nature* **184,** 60 (1959).
10. Mair, R. G. Response properties of rat olfactory bulb neurones. *J. Physiol. (Lond.)* **326,** 341–359 (1982).
11. Hinterwirth, A., Zeiner, R. & Tichy, H. Olfactory receptor cells on the cockroach antennae: responses to the direction and rate of change in food odour concentration. *Eur. J. Neurosci.* **19,** 3389–3392 (2004).
12. Liu, L. *et al. Drosophila* hygrosensation requires the TRP channels water witch and nanchung. *Nature* **450,** 294–298 (2007).
13. Miyawaki, A. *et al.* Fluorescent indicators for Ca^{2+} based on green fluorescent proteins and calmodulin. *Nature* **388,** 882–887 (1997).
14. Schiller, P. H. The ON and OFF channels of the visual system. *Trends Neurosci.* **15,** 86–92 (1992).
15. Richmond, J. E., Davis, W. S. & Jorgensen, E. M. UNC-13 is required for synaptic vesicle fusion in *C. elegans. Nature Neurosci.* **2,** 959–964 (1999).
16. Nonet, M. L., Saifee, O., Zhao, H., Rand, J. B. & Wei, L. Synaptic transmission deficits in *Caenorhabditis elegans* synaptobrevin mutants. *J. Neurosci.* **18,** 70–80 (1998).
17. Speese, S. *et al.* UNC-31 (CAPS) is required for dense-core vesicle but not synaptic vesicle exocytosis in *Caenorhabditis elegans. J. Neurosci.* **27,** 6150–6162 (2007).
18. Coburn, C. M. & Bargmann, C. I. A putative cyclic nucleotide-gated channel is required for sensory development and function in *C. elegans. Neuron* **17,** 695–706 (1996).
19. Komatsu, J., Mori, I., Rhee, J.-S., Akaike, N. & Ohshima, Y. Mutations in a cyclic neucleotide-gated channel lead to abnormal thermosensation and chemosesnation in *C. elegans. Neuron* **17,** 707–718 (1996).
20. Daniels, S. A., Ailion, M., Thomas, J. H. & Sengupta, P. *egl-4* acts through a transforming growth factor-beta/SMAD pathway in *Caenorhabditis elegans* to regulate multiple neuronal circuits in response to sensory cues. *Genetics* **156,** 123–141 (2000).
21. Hirose, T. *et al.* Cyclic GMP-dependent protein kinase EGL-4 controls body size and lifespan in *C. elegans. Development* **130,** 1089–1099 (2003).
22. L'Etoile, N. D. *et al.* The cyclic GMP-dependent protein kinase EGL-4 regulates olfactory adaptation in *C. elegans. Neuron* **36,** 1079–1089 (2002).
23. Tobin, D. *et al.* Combinatorial expression of TRPV channel proteins defines their sensory functions and subcellular localization in *C. elegans* neurons. *Neuron* **35,** 307–318 (2002).
24. Chalasani, S. H. *et al.* Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans. Nature* **450,** 63–70 (2007).
25. de Bruyne, M., Foster, K. & Carlson, J. R. Odor coding in the *Drosophila* antenna. *Neuron* **30,** 537–552 (2001).
26. Michel, W. C., McClintock, T. S. & Ache, B. W. Inhibition of lobster olfactory receptor cells by an odor-activated potassium conductance. *J. Neurophysiol.* **65,** 446–453 (1991).
27. Hills, T., Brockie, P. J. & Maricq, A. V. Dopamine and glutamate control area-restricted search behavior in *Caenorhabditis elegans. J. Neurosci.* **24,** 1217–1225 (2004).
28. Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans. Proc. Natl Acad. Sci. USA* **102,** 3184–3191 (2005).
29. Faumont, S., Miller, A. C. & Lockery, S. R. Chemosensory behavior of semi-restrained *Caenorhabditis elegans. J. Neurobiol.* **65,** 171–178 (2005).
30. Bargmann, C. I. & Avery, L. Laser killing of cells in *Caenorhabditis elegans. Methods Cell Biol.* **48,** 225–250 (1995).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the *Caenorhabditis* Genetics Center for strains, C. Frøkjær-Jensen for strain integration and D. Julius for the *TRPV1* complementary DNA. Support was provided by grants from the National Institutes of Health (MH051383 to S.R.L.; DA016445 to W.R.S.), National Science Foundation (IOB-0543643 to S.R.L.) and Human Frontier Science Program (to W.R.S.).

Author Contributions H.S. planned and performed the experiments first revealing the ASE ON–OFF function and the effects of transduction and synaptic mutants, made imaging and direct activation reagents, acquired supplementary ion-sensitivity data and drafted the manuscript. T.R.T. planned and performed ON–OFF, dose-response, genetics, direct activation and ablation experiments in the text and Supplementary Information, made imaging reagents, generated figures, and co-wrote the final manuscript. S.F. planned and performed ion selectivity and synaptic mutant imaging in the text and Supplementary Information. M.E. developed dose-response methodologies. S.R.L. planned imaging and behavioural experiments, devised the derivative model, and co-wrote the final manuscript. W.R.S. planned imaging and genetics experiments, and drafted the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to S.R.L. (shawn@uoregon.edu).

METHODS

Calcium imaging. Young adult animals were glued to a 1.5% agarose pad (30 mM *N*-tris(hydroxymethyl)methyl-3-aminopropanesulphonic acid (TAPS) buffer, pH 9) using cyanoacrylate glue (Nexaband S/C, Abbott Laboratories); animals were briefly cooled during the gluing process. A two-channel, gravity-fed perfusion pencil (360 μm tip), the flow of which was controlled by programmable upstream valves (AutoMate Scientific), was placed in close proximity (~0.5 mm) to the nose of each animal. The flow rate was 0.3 ml min^{-1} and the total bath volume was ~0.2 ml. Local perfusion, and thus rapid solution exchange ($t_{1/2} = 0.5$ s), was achieved by locating the worm in a fluid plume between the inflow perfusion pencil and the outflow pipette. Solutions contained the indicated amount of NaCl plus (in mM): HEPES-NaOH (10), pH 7.1, $CaCl_2$ (1) and $MgSO_4$ (1). Osmolarity was adjusted to 350 mOsm with glycerol. Optical recordings were performed on a Zeiss Axiovert 135, using a Zeiss Plan-Apochromat ×63 oil, 1.4 numerical aperture objective. The microscope was fitted with a Hamamatsu ORCA AG CCD camera (Hamamatsu Photonics), an Optical Insights Dual-View beam splitter (Photometrics), and an electronic shutter (ASI). Images were acquired at 10–20 Hz using MetaVue software (version 6.2r2, Molecular Devices). Image stacks were analysed using the Jamlyze program, as previously described[31]. The emission ratio was computed as (yellow fluorescent protein (YFP) intensity)/(cyan fluorescent protein (CFP) intensity) − 0.65, where the latter term corrects for CFP bleed-through into the YFP channel. Photobleaching was corrected by fitting a single exponential to inactive portions of the ratio trace and dividing the latter by the fitted curve. Imaging data in Supplementary Figs 2 and 5 were collected on a Zeiss Axioskop 2 upright microscope. In this setup, both ASE neurons were imaged simultaneously using a customized manipulator that rotated the worm such that both neurons were in the same focal plane. A constant flow of the pre-stimulus solution was delivered to the recording chamber at a rate of 1 ml min^{-1}. Stimuli were delivered by a perfusion pipette (100 μm tip), the position of which was controlled by a motorized stage (Polytec/PI M-111.1 DG microtranslation stage). During stimulation, the tip of the pipette was moved from downstream of the animal to within 50 μm of the worm's nose.

Selective activation of ASE neurons. The capsaicin imaging setup was identical to that used for NaCl imaging. The pre-stimulus saline was identical to the 40 mM NaCl imaging solution, except that it contained 0.25% ethanol. For the stimulus saline, a stock solution of capsaicin in ethanol was diluted in the pre-stimulus saline to 25 μM (ASEL experiments) or 5 μM (ASER experiments). We found that 25 μM capsaicin was required to evoke consistent ASEL calcium transients in the *gcy-7*::TRPV1 strain, whereas 5 μM was required to evoke consistent ASER calcium transients in the *gcy-5*::TRPV1 strain. The final ethanol concentration in the stimulus saline was always 0.25%. Behavioural responses to capsaicin were obtained using the previously described semi-restrained worm assay[29]. In brief, animals were affixed to a 1.5% agarose pad (30 mM TAPS, pH 9) using a small drop of cyanoacrylate glue placed near the terminal bulb of the pharynx, submerged in the pre-stimulus saline, and allowed to recover for 3–4 min. Under these conditions, semi-restrained animals exhibit behavioural correlates of forward and reverse locomotion as well as omega bends[29]. The behaviour of individual worms was scored manually by pressing computer keys to record changes in behavioural state. The strains, solutions and perfusion system were the same as those used in capsaicin imaging experiments. Experimental and control animals were tested in an alternating sequence, with the experimenter blind to the animal condition.

Neuronal ablations. Cells were ablated using a laser as described previously[30]. L1 larvae (N2 strain) were mounted on 2.5% agarose pads containing 5–7 mM of the immobilizing agent NaN_3. ASE neurons were identified by position. Animals were remounted 1–3 h after surgery to confirm the ablation. Animals with collateral damage were discarded. Sham-operated animals were treated in the same manner except that the laser was not fired. Step-response behaviour was tested three days after ablation.

Step-response assay. The step-response apparatus has been described[8]. Worms were placed individually on a thin porous membrane supported over a yoked pair of miniature inverted showerheads. Each showerhead emitted a saline solution with a different salt concentration. Stepwise concentration changes were delivered by sliding the showerhead assembly relative to the worm. The behaviour of individual worms was scored manually by pressing computer keys to record changes in behavioural state. Ablated and sham animals were tested in an alternating sequence, with the experimenter blind to the animal condition. Solutions used in the assay were identical to the solutions used during NaCl-imaging experiments. These solutions induced larger down-step behavioural responses than reported previously[8]; we determined that the increase in response was caused by the presence of 1 mM Ca^{2+} in the solutions used here (data not shown).

Behavioural statistics. The forward probabilities (P) for up-step experiments were transformed as 2 arcsin(P) to compensate for the compression of variance at extremes of the probability scale[32]; this transformation is conservative because it increases the variance at the extremes. Statistical significance was assessed by a two-factor repeated-measures ANOVA applied to forward probabilities during a post-stimulus analysis window. In addition, we made planned pairwise comparisons (t-tests) on mean behavioural-state probabilities at each time point in the analysis window. The significance level of each t-test was adjusted using the Bonferroni correction for multiple comparisons. To determine *a priori* the analysis window length to use in each ablation experiment condition (Fig. 4b–g and Supplementary Fig. 6), we identified the last post-stimulus time point at which there was a significant difference in post hoc comparisons (t-tests) between shams responding to a step and unstimulated control animals (a no-response group). The window extended from the first post-stimulus time point to this last significantly different time point. The list of analysis window lengths is as follows: 50 s for 9 mM up-step (Fig. 4b, d and f), 40 s for 9 mM down-step (Fig. 4c, e and g), 60 s for 40 mM up-step (Supplementary Fig. 6a, c and e) and 50 s for 40 mM down-step (Supplementary Fig. 6b, d and f).

31. Kerr, R. *et al.* Optical imaging of calcium transients in neurons and pharyngeal muscle of *C. elegans. Neuron* **26,** 583–594 (2000).
32. Winer, B. J., Brown, D. R. & Michels, K. M. *Statistica Principles in Experimental Design* 3rd ed. 354–358 (McGraw-Hill, Boston, 1991).

LETTERS

Formation of accumbens GluR2-lacking AMPA receptors mediates incubation of cocaine craving

Kelly L. Conrad[1], Kuei Y. Tseng[2], Jamie L. Uejima[3], Jeremy M. Reimers[1], Li-Jun Heng[2], Yavin Shaham[3], Michela Marinelli[2] & Marina E. Wolf[1]

Relapse to cocaine use after prolonged abstinence is an important clinical problem. This relapse is often induced by exposure to cues associated with cocaine use. To account for the persistent propensity for relapse, it has been suggested[1] that cue-induced cocaine craving increases over the first several weeks of abstinence and remains high for extended periods. We and others identified an analogous phenomenon in rats that was termed 'incubation of cocaine craving': time-dependent increases in cue-induced cocaine-seeking over the first months after withdrawal from self-administered cocaine[2–4]. Cocaine-seeking requires the activation of glutamate projections that excite receptors for α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA) in the nucleus accumbens[5–7]. Here we show that the number of synaptic AMPA receptors in the accumbens is increased after prolonged withdrawal from cocaine self-administration by the addition of new AMPA receptors lacking glutamate receptor 2 (GluR2). Furthermore, we show that these new receptors mediate the incubation of cocaine craving. Our results indicate that GluR2-lacking AMPA receptors could be a new target for drug development for the treatment of cocaine addiction. We propose that after prolonged withdrawal from cocaine, increased numbers of synaptic AMPA receptors combined with the higher conductance of GluR2-lacking AMPA receptors[8,9] causes increased reactivity of accumbens neurons to cocaine-related cues, leading to an intensification of drug craving and relapse.

For 10 days we trained rats for 6 h a day to nose-poke to receive intravenous cocaine or saline infusions (Fig. 1a, b); these infusions were paired with a 5-s light cue. After 1 or 45 days of withdrawal from self-administration of cocaine or saline, we assessed cue-induced cocaine-seeking in a 30-min extinction test. In this test, rats were exposed to cues previously associated with cocaine availability, but nose-poke responding in the previously active hole (a measure of cocaine-seeking) did not result in cocaine infusions. Consistent with prior results[2,3] was our observation that cue-induced cocaine-seeking was significantly greater on withdrawal day 45 than on withdrawal day 1 (Fig. 1c), confirming that cocaine craving incubates over time. On the basis of a critical role for glutamate-dependent plasticity in other addiction models[10–12], and our previous work[13,14], we proposed that time-dependent increases in accumbens AMPA receptor transmission underlie the incubation of cocaine craving. To test this hypothesis, we trained rats to self-administer cocaine or saline (as described above) for subsequent biochemical analysis after 1 or 45 days of withdrawal. The experimental groups were as follows: withdrawal day 1, saline (WD1-SAL); withdrawal day 1, cocaine (WD1-COC); withdrawal day 45, saline (WD45-SAL); and withdrawal day 45, cocaine (WD45-COC). We determined AMPA receptor distribution with a bis(sulphosuccinimidyl)suberate (BS^3) protein crosslinking assay that enables the quantification of surface and intracellular receptor pools in tissue harvested after treatments *in vivo*[13,14] (Supplementary Information).

We found substantial (2–3-fold) increases in surface, intracellular and total GluR1 levels in the WD45-COC group compared with all other groups (Fig. 2a–c), as well as a more modest increase in the surface/intracellular ratio of GluR1 (Fig. 2d). Thus, the major effect of prolonged withdrawal from cocaine is increased expression of GluR1, rather than redistribution of pre-existing GluR1, suggesting either increased GluR1 synthesis or decreased GluR1 degradation. The magnitude of increased GluR1 expression indicates a locus in medium spiny neurons, which make up more than 90% of accumbens neurons[15]. We found opposite changes in GluR1 on withdrawal day 1: rats in the WD1-COC group had significantly lower surface, intracellular and total GluR1 levels (Fig. 2a–c). Intermediate

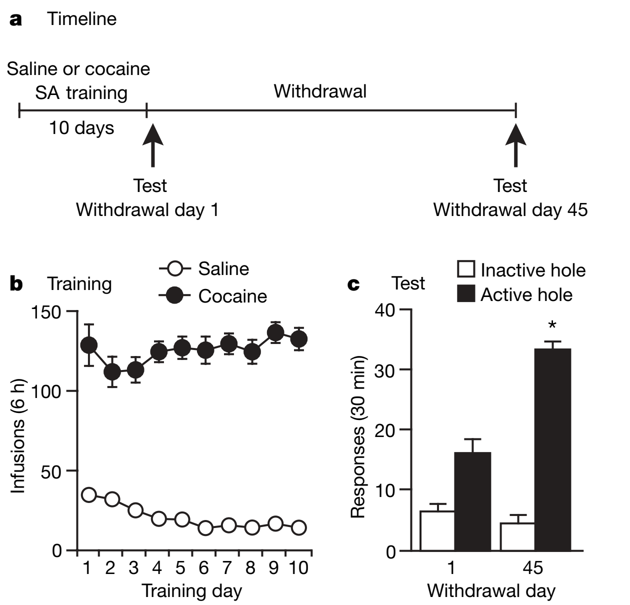


Figure 1 | Time-dependent increases in cue-induced cocaine-seeking (incubation of cocaine craving). **a**, Experimental timeline. SA, self-administration. **b**, Training: the number of infusions (each paired with a 5-s light cue) during training are shown (means ± s.e.m.). Cocaine (0.5 mg kg^{-1} per infusion) supported self-administration, as indicated by high nose-poking in the active hole. Responding in the inactive hole was very low (not shown). **c**, Drug-seeking tests: shown are the number (means and s.e.m.) of nose-pokes in the previously active hole (a measure of cocaine-seeking) and inactive hole during a 30-min test performed under extinction conditions (nose-pokes deliver cue but not cocaine). Cocaine-seeking increased on withdrawal day 45 (withdrawal day × hole interaction: $F_{1,12} = 14.9$, $P < 0.01$) ($n = 7$ per group). Asterisk, significantly different ($P < 0.05$) from withdrawal day 1.

[1]Department of Neuroscience and [2]Department of Cellular and Molecular Pharmacology, Rosalind Franklin University of Medicine and Science, 3333 Green Bay Road, North Chicago, Illinois 60064, USA. [3]Behavioral Neuroscience Branch, IRP/NIDA/NIH, 251 Bayview Boulevard, Baltimore, Maryland 21224, USA.

effects were observed in an additional experiment in which we examined receptor expression on withdrawal days 3 and 21 (Supplementary Fig. 1), indicating that GluR1 levels increase gradually after withdrawal from cocaine. For GluR2 we found a small increase in the surface/intracellular ratio in cocaine-exposed rats on withdrawal day 1, but no changes on withdrawal day 45 (Fig. 2h). For GluR3 we found increased surface expression on both withdrawal days 1 and 45, indicating a time-independent effect (Fig. 2i).

These results suggest that after prolonged withdrawal from cocaine, the normal complement of GluR2-containing AMPA receptors is supplemented by the addition of GluR2-lacking receptors (GluR1/3 and/or homomeric GluR1). We obtained additional support for this conclusion from a quantitative co-immunoprecipitation experiment (Supplementary Fig. 2). This effect is specific to accumbens AMPA receptors: we found no evidence for the formation of GluR2-lacking

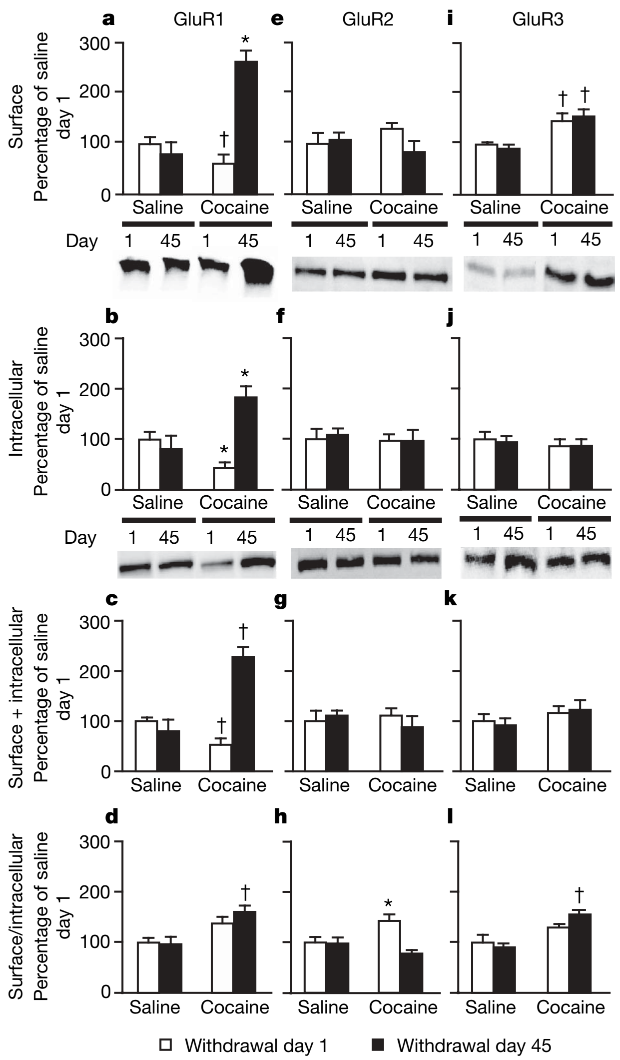


Figure 2 | Accumbens GluR1 and GluR3 expression increase after withdrawal from cocaine self-administration. a–d, GluR1 increased markedly in cocaine-exposed rats on withdrawal day 45 (surface (**a**), intracellular (**b**) and total (**c**) GluR1; drug exposure × day interaction, $F_{1,56} = 9.9$, $F_{1,56} = 9.2$ and $F_{1,56} = 12.1$, respectively; $P < 0.01$). **d**, Surface/intracellular ratio. **e–h**, GluR2 was unchanged except for a small increase in the surface/intracellular ratio in cocaine-exposed rats on withdrawal day 1 (drug exposure × day interaction, $F_{1,53} = 4.0$; $P < 0.01$). **i–l**, Surface GluR3 and the surface/intracellular ratio increased after self-administration of cocaine (drug exposure, $F_{1,48} = 4.4$ and $F_{1,48} = 3.9$, respectively; $P < 0.05$). Data (means and s.e.m., n = 12–18 per group) are expressed as percentages of the values in the saline group on withdrawal day 1. Asterisk, significantly different ($P < 0.05$) from other conditions; dagger, significantly different ($P < 0.05$) from saline group on withdrawal day 1.

AMPA receptors in the ventral tegmental area or cingulate cortex after withdrawal from cocaine (Supplementary Figs 3, 4), nor did we find significant changes in accumbens N-methyl-D-aspartate receptor subunits (Supplementary Fig. 5).

The accumbens consists of two major subregions, termed core and shell, which can be distinguished on the basis of connectivity and morphology[15]. The core and shell have different roles in drug-related behaviours, with some evidence suggesting that the core is more significant in cue-induced cocaine-seeking[16]. To study potential core-shell differences, we assessed another cohort of cocaine self-administering rats after 1 or 45 days of withdrawal. We divided the accumbens into core and shell subregions, crosslinked with BS^3, and analysed GluR1–3 (Supplementary Fig. 6). In the core we found robust time-dependent increases in GluR1 and modest increases in GluR3; in the shell we found that surface GluR1 was increased on withdrawal day 45. These results suggest that GluR2-lacking AMPA receptors form in both core and shell, but this effect may be more pronounced in the core.

Next we determined whether the time-dependent changes in AMPA receptor expression were influenced by performing a test for cue-induced cocaine-seeking (under extinction conditions). We trained rats to self-administer cocaine as described above. We assessed the brains of four groups of rats that were either tested ('test') or not tested ('no-test') for cue-induced cocaine-seeking after 1 or 45 days of withdrawal from cocaine; rats in the test condition were killed immediately after the 30-min cocaine-seeking test. We found increased surface and total GluR1 levels on withdrawal day 45 (Supplementary Fig. 7a–c) in both the test and no-test conditions, replicating results from our first experiment (Fig. 2a–c). No-test rats also showed a small decrease in the surface/intracellular ratio of GluR2 on withdrawal day 45 (Supplementary Fig. 7h). These data suggest that the test for cocaine-seeking had a minimal effect on AMPA receptor distribution in the accumbens.

To confirm our biochemical results we performed whole-cell patch-clamp recordings of medium spiny neurons in the accumbens core after 42–47 days of withdrawal from saline or cocaine self-administration. GluR2-lacking AMPA receptors have unique properties: permeability to Ca^{2+}, resulting in greater conductance, and inwardly rectifying currents due to voltage-dependent block by polyamines[8,9]. Current–voltage relationships of evoked excitatory postsynaptic currents (EPSCs; Supplementary Information) in accumbens neurons revealed significantly greater inward rectification in the cocaine-exposed group (Fig. 3a–c). Furthermore, bath application of 1-naphthylacetylsperimine (Naspm), a selective blocker of GluR2-lacking AMPA receptors, decreased evoked EPSC amplitude only in neurons recorded from the cocaine-exposed group (Fig. 3d–g). Thus, GluR2-lacking AMPA receptors contribute significantly to accumbens synaptic transmission only after prolonged withdrawal from cocaine. In addition, we found that neurons from cocaine-exposed rats showed a change in the distribution of spontaneous EPSC (sEPSC) amplitude as a result of an increased number of high-amplitude sEPSC (Supplementary Fig. 8b). Both the results with Naspm and the increased sEPSC amplitude predict enhanced responsiveness of accumbens neurons to excitatory inputs after prolonged withdrawal from cocaine. Neurons from cocaine-exposed rats also showed an increased frequency of AMPA-receptor-mediated sEPSCs (Supplementary Fig. 8a). This is unlikely to reflect increased probability of release, because the paired-pulse ratio did not differ between cocaine-exposed and saline-exposed rats (Supplementary Fig. 8c, d). Increased sEPSC frequency may be due to the formation of new synaptic contacts in the accumbens after withdrawal from cocaine[17].

To test the functional role of new GluR2-lacking receptors, we injected Naspm (or vehicle) into the accumbens of cocaine-exposed rats before tests for cue-induced cocaine-seeking. Naspm decreased cue-induced cocaine-seeking on withdrawal day 45, demonstrating that GluR2-lacking AMPA receptors mediate the expression of incubation of cocaine craving (Fig. 4a). Naspm did not alter cue-induced cocaine-seeking on withdrawal day 1 (Fig. 4a). This finding is consistent with a

lack of differences in AMPA receptor subunit expression and distribution on cocaine withdrawal day 1 compared with the drug-naive saline condition (Fig. 2), in which GluR2-lacking receptors are expressed at very low levels and contribute minimally to synaptic transmission in the accumbens (Fig. 3 and Supplementary Fig. 2). Naspm did not alter stable cocaine or sucrose self-administration (Fig. 4b).

We propose that the synaptic incorporation of GluR2-lacking AMPA receptors enhances the responsiveness of accumbens neurons to glutamate inputs from cortical and limbic regions, as a result of increases in the absolute number of surface AMPA receptors (Fig. 2) as well as the higher conductance of GluR2-lacking AMPA receptors[8,9]. Thus, when cocaine-associated cues are presented after prolonged withdrawal from cocaine and glutamate is released in the accumbens, neurons in this brain area respond more robustly, leading to enhanced cocaine-seeking.

Our results are consistent with a large body of literature implicating increased AMPA receptor transmission in the accumbens in cocaine-seeking[5–7] and psychomotor sensitization[13,14,18–21] after prolonged withdrawal from cocaine, and with the finding that increased neuronal activity in the accumbens correlates with the incubation of cocaine craving[22]. However, our results differ from the finding that cue-induced cocaine-seeking after prolonged withdrawal is not decreased in GluR1-knockout mice[23]. These latter results should be interpreted with caution in view of the potential for compensation during development and/or offsetting changes in other neuronal pathways. A previous study found that viral overexpression of GluR1 or GluR2 in the accumbens shell decreases cocaine-seeking during early withdrawal[24]. Many differences exist between our study and this previous study, including focus on core versus shell, long versus short withdrawal, and single versus multiple extinction tests. An important consideration is that our conclusions are based on measuring and manipulating endogenous surface AMPA receptors.

Recent work has highlighted the importance of GluR2-lacking AMPA receptors in long-term potentiation (LTP) and depression (LTD), experience-dependent plasticity, and synaptic scaling[8,9]. Synaptic scaling is a form of homeostatic plasticity in which prolonged activity blockade causes enhanced excitatory synaptic transmission. Synaptic scaling may have parallels to our model. After withdrawal from cocaine, cortical areas providing excitatory input to the accumbens show metabolic hypoactivity[10,25], raising the possibility that accumbens GluR2-lacking AMPA receptors scale up as a homeostatic response to prolonged decreases in synaptic activation. Scaling-induced increases in GluR1 have been reported to occur through increased dendritic GluR1 synthesis[26] as well as increased GluR1 protein stability[27].

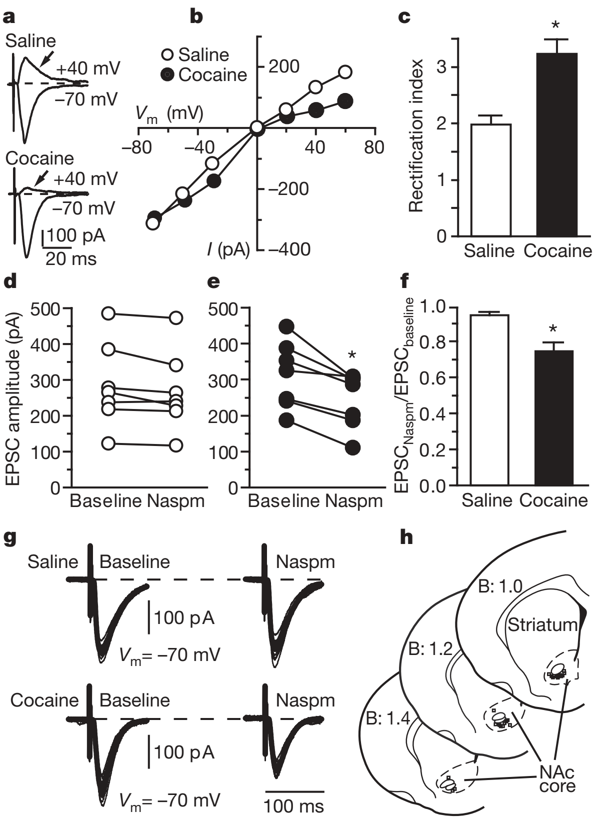


Figure 3 | GluR2-lacking AMPA receptors are detected in accumbens neurons after prolonged withdrawal from self-administration of cocaine. **a**, Evoked EPSC recorded after 42–47 days of withdrawal from saline or cocaine self-administration. **b**, Current–voltage relationships for neurons shown in **a**. **c**, Rectification index ($EPSC_{-70\,mV}/EPSC_{+40\,mV}$; Supplementary Information) from 13 and 8 neurons recorded from four cocaine-exposed and three saline-exposed rats ($t_{19} = 3.47$; asterisk, $P < 0.01$). Data are means and s.e.m. **d**, **e**, Naspm (200 μM, 5–10 min) decreased evoked EPSC amplitude in cocaine-exposed rats (**e**) ($t_6 = 4.72$; asterisk, $P < 0.01$, baseline versus Naspm, seven cells per group) but not in saline-exposed rats (**d**). **f**, The effect of Naspm illustrated as the evoked EPSC amplitude normalized to baseline ($t_{12} = 3.73$; asterisk, $P < 0.01$). **g**, Representative traces illustrating the effect of Naspm after 10 min of bath application. **h**, Location of recordings. Numbers indicate distance rostral to Bregma (B) in mm.

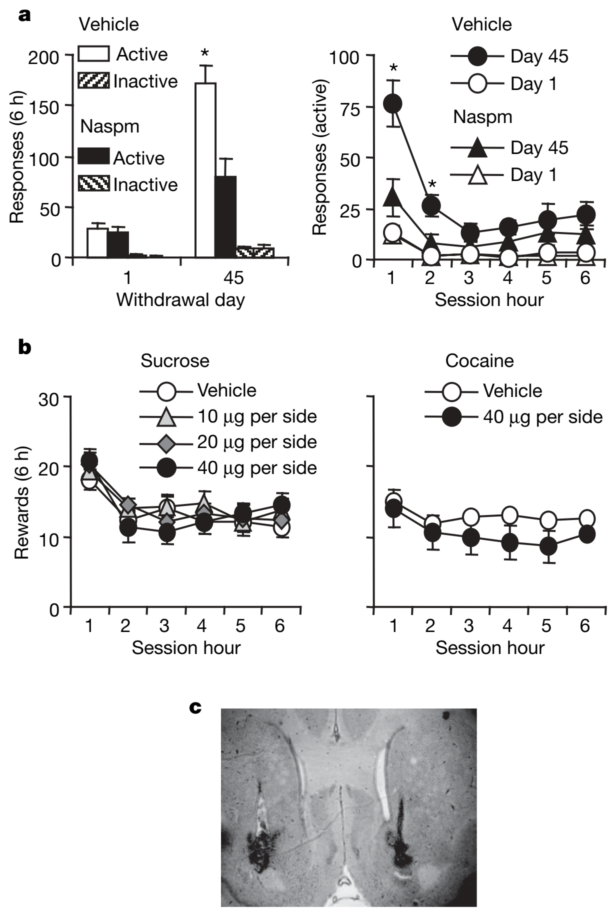


Figure 4 | Enhanced cue-induced cocaine-seeking after prolonged withdrawal from cocaine self-administration is inhibited by blockade of GluR2-lacking AMPA receptors. **a**, Cue-induced cocaine-seeking tests. Left: responses (means and s.e.m.) on the previously active or inactive levers after injections of Naspm or vehicle into the accumbens 15 min before extinction tests on withdrawal days 1 or 45 ($n = 10$–14 per group). Right: responses on previously active lever at each hour of test (Naspm dose × withdrawal day × session hour × lever interaction, $F_{1,45} = 4.6$; asterisk, $P < 0.05$ compared with other groups). **b**, Injections of Naspm into the accumbens had no effect on stable self-administration of oral sucrose or intravenous cocaine. Results are numbers of oral sucrose deliveries (left; 0.75 ml per delivery, $n = 10$) or intravenous cocaine deliveries (right; 0.75 mg kg^{-1} per infusion, $n = 5$) (means ± s.e.m.). **c**, Placement of cannulae, showing injector tips.

Here we have demonstrated that GluR2-lacking AMPA receptors are produced in the accumbens during prolonged withdrawal from cocaine and mediate the incubation of cocaine craving. Our work adds to a growing consensus that perturbations in synaptic transmission during disease states cause compensatory changes in AMPA receptor subunit composition that alter the properties of neuronal networks[8,9]. For cocaine addiction, the production of GluR2-lacking AMPA receptors may exacerbate disease processes by increasing the reactivity of accumbens neurons to cocaine-associated cues that promote craving and relapse. A question for future research is whether GluR2-lacking receptors in the accumbens also contribute to drug-induced and stress-induced cocaine craving and relapse that also occur after prolonged withdrawal[3,10]. Finally, our results, and those of others on the formation of GluR2-lacking AMPA receptors in the ventral tegmental area after acute exposure to cocaine[28,29], suggest that these receptors could be a new drug target for the treatment of addiction.

METHODS SUMMARY

All procedures are based on our previous work[2,3,13,14,30].

Behavioural procedures. Male rats were trained to nose-poke (biochemical and electrophysiology experiments) or lever-press (Naspm behavioural experiment) for 6 h a day for 10–12 days; each cocaine infusion was paired with a tone–light or light cue. After self-administration training, the rats were tested for cue-induced cocaine-seeking after 1 or 45 days of withdrawal. During testing, responding with a lever-press or a nose-poke led to contingent presentations of the cue previously paired with cocaine infusions, but not cocaine. Responding on the previously active lever or hole was the operational measure of cocaine-seeking.

Biochemistry. After the appropriate withdrawal period (or immediately after the drug-seeking test in Supplementary Fig. 7), the rats were decapitated. The accumbens was rapidly dissected and brain slices (400 μm thick) prepared with a tissue chopper. Crosslinking of slices with BS^3 (30 min), subsequent tissue processing, and quantification of surface and intracellular protein levels by SDS–PAGE and western blotting were performed as described in Methods. Values for surface, intracellular and total receptor subunit levels were normalized to total protein in the lane determined with Ponceau S.

Electrophysiology. Coronal slices (300 μm thick) containing the accumbens were obtained after 42–47 days of withdrawal from self-administration of saline or cocaine. Recordings were conducted in voltage-clamp configuration at 33–35 °C with patch electrodes filled with caesium gluconate, spermine (0.1 mM) and QX-314 (1 mM). Synaptic responses of medium spiny neurons were elicited by local stimulation of excitatory inputs with a bipolar electrode. Stimulation intensity (0.05–0.3 mA) was based on the minimum amount of current necessary to elicit a synaptic response with less than 15% variability in amplitude 10 min after obtaining the whole-cell configuration. Both spontaneous and evoked EPSCs were collected before and after bath application of Naspm (100–200 μM) for 10 min.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 December 2007; accepted 7 April 2008.
Published online 25 May 2008; corrected 3 July 2008 (details online).

1. Gawin, F. H. & Kleber, H. D. Abstinence symptomatology and psychiatric diagnosis in cocaine abusers. Clinical observations. *Arch. Gen. Psychiatry* **43,** 107–113 (1986).
2. Grimm, J. W., Hope, B. T., Wise, R. A. & Shaham, Y. Incubation of cocaine craving after withdrawal. *Nature* **412,** 141–142 (2001).
3. Lu, L., Grimm, J. W., Hope, B. T. & Shaham, Y. Incubation of cocaine craving after withdrawal: a review of preclinical data. *Neuropharmacology* **47** (Suppl 1), 214–226 (2004).
4. Neisewander, J. L. *et al.* Fos protein expression and cocaine-seeking behavior in rats after exposure to a cocaine self-administration environment. *J. Neurosci.* **20,** 798–805 (2000).
5. Cornish, J. L. & Kalivas, P. W. Glutamate transmission in the nucleus accumbens mediates relapse in cocaine addiction. *J. Neurosci.* **20,** RC89 (2000).
6. Di Ciano, P., Cardinal, R. N., Cowell, R. A., Little, S. J. & Everitt, B. J. Differential involvement of NMDA, AMPA/kainate, and dopamine receptors in the nucleus accumbens core in the acquisition and performance of pavlovian approach behavior. *J. Neurosci.* **21,** 9471–9477 (2001).
7. Suto, N. *et al.* Previous exposure to psychostimulants enhances the reinstatement of cocaine seeking by nucleus accumbens AMPA. *Neuropsychopharmacology* **29,** 2149–2159 (2004).
8. Cull-Candy, S., Kelly, L. & Farrant, M. Regulation of Ca^{2+}-permeable AMPA receptors: synaptic plasticity and beyond. *Curr. Opin. Neurobiol.* **16,** 288–297 (2006).
9. Liu, S. J. & Zukin, R. S. Ca^{2+}-permeable AMPA receptors in synaptic plasticity and neuronal death. *Trends Neurosci.* **30,** 126–134 (2007).
10. Kalivas, P. W. & Volkow, N. D. The neural basis of addiction: a pathology of motivation and choice. *Am. J. Psychiatry* **162,** 1403–1413 (2005).
11. Kauer, J. A. & Malenka, R. C. Synaptic plasticity and addiction. *Nature Rev. Neurosci.* **8,** 844–858 (2007).
12. Wolf, M. E. The role of excitatory amino acids in behavioral sensitization to psychomotor stimulants. *Prog. Neurobiol.* **54,** 679–720 (1998).
13. Boudreau, A. C. & Wolf, M. E. Behavioral sensitization to cocaine is associated with increased AMPA receptor surface expression in the nucleus accumbens. *J. Neurosci.* **25,** 9144–9151 (2005).
14. Boudreau, A. C., Reimers, J. M., Milovanovic, M. & Wolf, M. E. Cell surface AMPA receptors in the rat nucleus accumbens increase during cocaine withdrawal but internalize after cocaine challenge in association with altered activation of mitogen-activated protein kinases. *J. Neurosci.* **27,** 10621–10635 (2007).
15. Meredith, G. E. The synaptic framework for chemical signaling in nucleus accumbens. *Ann. NY Acad. Sci.* **877,** 140–156 (1999).
16. Everitt, B. J. & Robbins, T. W. Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. *Nature Neurosci.* **8,** 1481–1489 (2005).
17. Robinson, T. E. & Kolb, B. Structural plasticity associated with exposure to drugs of abuse. *Neuropharmacology* **47** (Suppl. 1), 33–46 (2004).
18. Goto, Y. & Grace, A. A. Dopamine-dependent interactions between limbic and prefrontal cortical plasticity in the nucleus accumbens: disruption by cocaine sensitization. *Neuron* **47,** 255–266 (2005).
19. Kourrich, S., Rothwell, P. E., Klug, J. R. & Thomas, M. J. Cocaine experience controls bidirectional synaptic plasticity in the nucleus accumbens. *J. Neurosci.* **27,** 7921–7928 (2007).
20. Pierce, R. C., Bell, K., Duffy, P. & Kalivas, P. W. Repeated cocaine augments excitatory amino acid transmission in the nucleus accumbens only in rats having developed behavioral sensitization. *J. Neurosci.* **16,** 1550–1560 (1996).
21. Yao, W. D. *et al.* Identification of PSD-95 as a regulator of dopamine-mediated synaptic and behavioral plasticity. *Neuron* **41,** 625–638 (2004).
22. Hollander, J. A. & Carelli, R. M. Cocaine-associated stimuli increase cocaine seeking and activate accumbens core neurons after abstinence. *J. Neurosci.* **27,** 3535–3539 (2007).
23. Mead, A. N., Zamanillo, D., Becker, N. & Stephens, D. N. AMPA-receptor GluR1 subunits are involved in the control over behavior by cocaine-paired cues. *Neuropsychopharmacology* **32,** 343–353 (2007).
24. Sutton, M. A. *et al.* Extinction-induced upregulation in AMPA receptors reduces cocaine-seeking behaviour. *Nature* **421,** 70–75 (2003).
25. Beveridge, T. J., Smith, H. R., Daunais, J. B., Nader, M. A. & Porrino, L. J. Chronic cocaine self-administration is associated with altered functional activity in the temporal lobes of non human primates. *Eur. J. Neurosci.* **23,** 3109–3118 (2006).
26. Ju, W. *et al.* Activity-dependent regulation of dendritic synthesis and trafficking of AMPA receptors. *Nature Neurosci.* **7,** 244–253 (2004).
27. O'Brien, R. J. *et al.* Activity-dependent modulation of synaptic AMPA receptor accumulation. *Neuron* **21,** 1067–1078 (1998).
28. Bellone, C. & Luscher, C. Cocaine triggered AMPA receptor redistribution is reversed *in vivo* by mGluR-dependent long-term depression. *Nature Neurosci.* **9,** 636–641 (2006).
29. Mameli, M., Balland, B., Lujan, R. & Luscher, C. Rapid synthesis and synaptic insertion of GluR2 for mGluR-LTD in the ventral tegmental area. *Science* **317,** 530–533 (2007).
30. Tseng, K. Y. & O'Donnell, P. Dopamine-glutamate interactions controlling prefrontal cortical pyramidal cell excitability involve multiple signaling mechanisms. *J. Neurosci.* **24,** 5131–5139 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. J. Wenthold for help in establishing quantitative co-immunoprecipitation methods used in Supplementary Fig. 2. This work was supported by US Public Health Service grants DA09621, DA015835, DA00453 and NARSAD (M.E.W.), DA020654 (M.M.), predoctoral National Research Service Award DA021488 (K.L.C.), Rosalind Franklin University of Medicine and Science start-up funds (K.Y.T.) and the Intramural Research Program of the National Institute on Drug Abuse (Y.S.).

Author Contributions K.L.C., M.M. and M.E.W. were responsible for overall study design. K.L.C. conducted and analysed cocaine self-administration experiments (except the Naspm experiment) and protein crosslinking studies. M.M. trained K.L.C. in drug self-administration procedures and helped with these experiments. J.M.R. conducted and analysed co-immunoprecipitation experiments. K.Y.T. designed electrophysiological experiments, L.J.H. conducted them, and K.Y.T. analysed the data. Y.S. and K.L.C. designed the Naspm behavioural experiment, J.L.U. performed it, and Y.S. and J.L.U. analysed the data. K.L.C., Y.S. and M.E.W. wrote the paper with the help of the other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.E.W. (marina.wolf@rosalindfranklin.edu).

METHODS

Subjects. The subjects were male Sprague–Dawley rats (Harlan; biochemical and electrophysiological experiments) and Long Evans rats (Charles River; Naspm behavioural experiment performed at the Intramural Research Program of the National Institute on Drug Abuse) weighing 250–275 g on arrival. The rats were housed individually on a reverse 12 h/12 h light–dark cycle (lights out at 09:00). Rats had access to water and food *ad libitum* at all times unless specified. All procedures followed NIH guidelines and were approved by the local Animal Care and Use Committees. Rats were implanted with either intravenous catheters or intravenous catheters plus bilateral cannulae aimed at the nucleus accumbens before saline or cocaine self-administration training (see below). The inclusion of saline-exposed control rats that were drug-free but still exposed to the same experimental conditions was used to control for effects of ageing on our molecular measures.

Surgical procedures. The rats were anaesthetized with isoflurane gas (Henry Schein), and flunixin meglumine was administered before surgery ($2\,\mathrm{mg\,kg^{-1}}$, intraperitoneally) as an analgesic. A silastic catheter was inserted into the right auricle through the external jugular vein, passed under the skin and fixed in the mid-scapular region. The rats recovered from surgery for at least seven days before beginning self-administration training sessions. During this time, catheters were flushed every 24–48 h with sterile 0.9% saline. The rats undergoing intravenous self-administration experiments together with intracranial infusions (Naspm behavioural experiment) were anaesthetized with sodium pentobarbital and chloral hydrate (60 and $25\,\mathrm{mg\,kg^{-1}}$, intraperitoneally), and permanent guide cannulae (23-gauge; Plastics One) were implanted bilaterally 1 mm above the nucleus accumbens and were aimed at the core subregion (coordinates from Bregma: 6° angle aimed medially; anteroposterior, +1.7 mm; mediolateral, ± 2.5 mm; dorsoventral, −6.0 mm)[31]. After implantation of cannulae, silastic catheters were inserted into the jugular vein, attached to a modified 22-gauge cannula and mounted to the rat's skull with dental cement (see refs 32, 33). Buprenorphine ($0.1\,\mathrm{mg\,kg^{-1}}$, subcutaneously) was given after surgery as an analgesic, and the rats recovered for 7–10 days before behavioural testing began. During the recovery and training phases for these rats, catheters were flushed every 24–48 h with sterile 0.9% saline plus the antibiotic gentamicin ($0.08\,\mathrm{mg\,ml^{-1}}$).

Intracranial injections. The intracranial injection methods were based on our previous studies[32,33]. Naspm (Sigma-Aldrich) was dissolved in phosphate-buffered saline (PBS). Injections of vehicle or Naspm (10, 20 and 40 μg per side) were made with Hamilton syringes (Hamilton) that were connected to 30-gauge injectors (Plastics One). A volume of 0.5 μl was infused into each side over the course of 1 min and the injector was left in place for 1 min after the injections. The rats were tested within 15 min after intracranial injections. The doses of Naspm were based on previous reports[34,35] and on an initial study with sucrose-trained rats (see below). At the end of the experiments, the rats were deeply anaesthetized, their brains were removed, and coronal sections (40 μm) were sliced on a cryostat and stained with cresyl violet (ICN Biomedicals Inc.). Placements of cannulae were verified under a microscope and their anatomical location is shown in Fig. 4c.

Cocaine self-administration training. After recovery from surgery, the rats were allowed to self-administer cocaine or saline for 6 h a day for 10 days (biochemical and electrophysiological experiments) or for 10–12 days (Naspm behavioural experiment). The self-administration chambers (MED Associates) were located in sound-attenuating cabinets. Rats were either housed chronically in these chambers (for the Naspm experiments) or placed daily in these chambers; sessions began approximately at the start of the dark cycle. For the Naspm experiments, the self-administration chambers were equipped with two levers. Presses on one (active, retractable) activated the infusion pump and delivered an infusion of cocaine ($0.75\,\mathrm{mg\,kg^{-1}}$); presses on the other (inactive, stationery) had no effects. A fixed-ratio-1 reinforcement schedule (each lever press is reinforced with one cocaine infusion) was used, with a 40-s timeout period after each infusion; cocaine infusions were accompanied by a 5-s tone–light cue. Each session began with the insertion of the active lever and the illumination of a houselight that remained on for the entire session. At the end of each session, the houselight was turned off and the active lever was retracted. To facilitate the acquisition of cocaine self-administration, food was removed from the chambers during the 6-h sessions of the first three training days. The number of cocaine infusions was limited to 20 per hour to prevent overdose. For all other experiments, the self-administration chambers were equipped with two holes located 2 cm above the floor. Nose-poking in the active hole activated the infusion pump and delivered an infusion of saline or cocaine ($0.5\,\mathrm{mg\,kg^{-1}}$); nose-poking in the inactive hole had no consequences. In addition to activating the infusion pump, nose-poking in the active hole was paired with a 5-s discrete light cue, located inside the nose hole. A time-out period of 10 s was used during the first hour or for the first ten infusions (whichever occurred first) and then the time-out period was extended to 30 s for the remaining hours, to prevent cocaine overdose. Food and water were present at all times. For sucrose self-administration (results shown in Fig. 4b; $n = 10$), procedures were the same as those described above for cocaine self-administration for the Naspm experiment, except that active lever presses led to the delivery of 0.75 ml of 10% sucrose solution into receptacles located near the lever. After stable sucrose self-administration behaviour was achieved, the rats were injected every other day with vehicle or Naspm (10, 20 or 40 μg per side) into the accumbens 15 min before the test sessions, which were separated by regular training days. The order of the injections of the vehicle and the different Naspm doses was counterbalanced. Naspm (40 μg per side) or its vehicle was also injected during cocaine self-administration in a subgroup of rats ($n = 5$), as described above for sucrose.

Tests for cue-induced cocaine-seeking. At the end of the training phase, the rats were returned to the animal facility, where they remained for 1 or 45 days (the rats in the late withdrawal period were handled several times per week). After this time, they were brought back to the self-administration chambers, where they were tested for cue-induced cocaine-seeking under extinction conditions; that is, all conditions were the same as during training, with the exception that responding on the active device was not reinforced with drug. During the extinction tests, lever or nose-poke responses led to contingent presentations of the tone–light or light cue previously paired with cocaine infusions. The number of responses in the previously active lever or hole was used as a measure of cocaine-seeking.

Protein crosslinking. Each experimental group consisted of 7–18 rats, with the exception of cocaine withdrawal day 21 ($n = 5$). The rats were decapitated, their brains were rapidly removed, and the nucleus accumbens (or other region of interest) was dissected on ice from a 2-mm coronal section obtained with a brain matrix. Tissue was immediately chopped into 400-μm slices with a McIllwain tissue chopper (Vibratome). Slices were added to Eppendorf tubes containing ice-cold artificial cerebrospinal fluid (aCSF), which was spiked with 2 mM BS^3 (Pierce Biotechnology) immediately after addition of the tissue. Slices were crosslinked for 30 min at 4 °C with gentle agitation. Crosslinking was terminated by the addition of 100 mM glycine (10 min at 4 °C). Slices were pelleted by brief centrifugation, resuspended in ice-cold lysis buffer containing protease and phosphatase inhibitors, sonciated for 5 s to disrupt tissue, and centrifuged (20,800*g* for 2 min) as described previously[13,14]. The supernatant fraction was divided into aliquots and stored at −80 °C. BS^3 is a membrane-impermeant crosslinking agent; it therefore selectively crosslinks cell-surface proteins, forming high-molecular-mass aggregates. Intracellular proteins are not modified and thus retain their normal molecular mass. This enables surface and intracellular pools of a particular protein to be distinguished by SDS–PAGE and western blotting. Variants of this assay have been used previously to measure the surface expression of glutamate receptors in dissociated cells and brain slices[36–44]. We adapted the assay to detect receptor redistribution produced after treatments *in vivo*[13]. We and others have shown that incubation of brain slices or dissociated cultures with BS^3 does not crosslink intracellular proteins (for example actin, tubulin, synapsin, tyrosine hydroxylase and protein kinases) unless BS^3 crosslinking is performed in a lysed preparation[13,14,36–40,42,43].

Western blot analysis of glutamate receptor subunits in crosslinked tissue. Samples were run on 4–15% gradient Tris-HCl gels (Bio-Rad) under reducing conditions, proteins were transferred to poly(vinylidene difluoride) (PVDF) membranes, and membranes were washed in Tris-buffered saline (TBS) and blocked with 1% goat serum plus 5% nonfat dried milk in TBS containing Tween 20 (TBS-T). Membranes were incubated overnight at 4 °C with the following primary antibodies: GluR1 (1:500 dilution; Millipore), GluR2 (1:1,000 dilution, Millipore), GluR3 (1:500 dilution; Millipore), NR1 (1:500 dilution; Millipore), NR2A (1:2,000 dilution, Santa Cruz Biotechnology), and NR2B (1:1,000 dilution; Millipore). Not all lots of the NR1 antibody gave satisfactory results in this assay. Membranes were washed with TBS-T solution, incubated for 60 min with horseradish peroxidase (HRP)-conjugated anti-rabbit IgG or anti-mouse IgG (1:10,000 dilution; Upstate Biotechnology), washed with TBS-T, rinsed with double-deionized water, and immersed in enhanced chemiluminescence (ECL)-detecting substrate (Amersham GE). Images were captured with Versa Doc Imaging Software (Bio-Rad). Diffuse densities of surface and intracellular bands were determined with Quantity One software (Bio-Rad). Values for surface, intracellular and total (surface plus intracellular) protein levels were normalized to total lane protein determined with Ponceau S (Sigma-Aldrich) and analysed with TotalLab (Nonlinear Dynamics). The surface/intracellular ratio did not require normalization because both values were determined in the same lane.

Quantitative co-immunoprecipitation. Using the methods developed in refs 45, 46 and with the help of advice from the Wenthold laboratory, we quantitatively determined AMPA receptor subunit composition in the accumbens after 45 days of withdrawal from cocaine or saline self-administration. In brief, the

rats were decapitated, their brains were rapidly removed, and the accumbens was dissected on ice from a 2-mm coronal section obtained with a brain matrix. Tissue from three rats from each experimental group was combined and homogenized in 50 mM Tris-HCl pH 7.4 containing protease inhibitor cocktail (Calbiochem) (40–60 mg wet weight ml^{-1}). The membranes were sedimented by centrifugation at 100,000*g* for 30 min at 4 °C. The pellet was then solubilized at 37 °C for 45 min with 1% Triton X-100 in 50 mM Tris-HCl pH 7.4 containing 1 mM EDTA. Insoluble material was removed by centrifugation at 100,000*g* for 30 min at 4 °C. The supernatant was stored at −80 °C until use. For co-immunoprecipitation, 3–5 μg of antibody (GluR1, GluR2, GluR2/3 or GluR4) or an equal amount of control IgG was incubated with 10–20 μl of 50% Protein A-agarose slurry (Pierce) for 4 h at 4 °C. The pellet was collected by centrifugation at 1,000*g* for 30 s and washed three times with TBS containing 0.1% Triton X-100. Membrane preparation (100 μl) was incubated overnight with the washed pellet at 4 °C. The agarose-bound antibody was pelleted by centrifugation at 1,000*g* for 30 s. This created two fractions, the bound (pellet) and unbound (supernatant). The unbound fraction was then subjected to another round of immunoprecipitation. Two rounds of immunoprecipitation pulled down more than 95% of the target AMPA receptor subunit (for example, in Supplementary Fig. 2, after immunoprecipitation for GluR1, no GluR1 is detected in the unbound fraction by immunoblotting). After the final immunoprecipitation, the unbound fraction was mixed with an equal volume of sample treatment buffer (Invitrogen) and heated to 70 °C for 10 min. For western blot analysis, samples were run on 4–12% Bis-Tris gels (Invitrogen) and transferred to PVDF membranes for immunoblotting. Membranes were washed in double-deionized water and blocked with 1% goat serum plus 5% Carnation milk in 0.05% Tween 20 in TBS, pH 7.4 for 1 h at 20–23.5 °C. Membranes were then incubated overnight with subunit-specific antibodies (GluR1, 1:500 dilution; GluR2/3, 1:2,000 dilution; GluR2, 1:1,000 dilution; GluR3, 1:500 dilution; all from Millipore) at 4 °C. We did not immunoprobe for GluR4 because it is not present in medium spiny neurons[47,48]; consistent with this, after immunoprecipitation for GluR4, 100% of GluR1, GluR2 and GluR3 is detected in the unbound fraction (Supplementary Fig. 2). Membranes were then washed with TBS-Tween solution, incubated for 60 min with HRP-conjugated anti-rabbit IgG or anti-mouse IgG (1:10,000 dilution; Upstate Biotechnology), and washed again with TBS-Tween, followed by TBS. Membranes were then rinsed with double-deionized water, immersed in ECL-detecting substrate (Amersham GE) for 1 min, and detected with VersaDoc imaging software (Bio-Rad) (between 5 and 60 s, depending on the antibody). Diffuse densities of bands were determined with Quantity One software (Bio-Rad). The percentage of total AMPA receptor subunit remaining in the unbound fraction was calculated on the basis of the standard curve created from control IgG immunoprecipitated tissue, as described in the legend to Supplementary Fig. 2.

Electrophysiology. As reported previously[30], the rats were anaesthetized with chloral hydrate (400 mg kg^{-1}, intraperitoneally) before being decapitated. Brains were rapidly removed into ice-cold aCSF containing (in mM): 125 NaCl, 25 $NaHCO_3$, 12.5 glucose, 3.5 KCl, 1.25 NaH_2PO_4, 0.5 $CaCl_2$, 3 $MgCl_2$, 0.05 D-2-amino-5-phosphonovalerate, and 0.05 picrotoxin (pH 7.45, 295–305 m-osM). Coronal slices (300 μm thick) containing the nucleus accumbens were cut in ice-cold aCSF with a Vibratome, and incubated in warm (about 35 °C) aCSF solution constantly oxygenated with 95% O_2/5% CO_2 for at least 60 min before recording. During recordings, aCSF was delivered at 2 ml min^{-1}, the concentration of $CaCl_2$ was increased to 2 mM, and the concentration of $MgCl_2$ was decreased to 1 mM. Patch pipettes (6–9 MΩ) were pulled from 1.5-mm borosilicate glass capillaries (WPI) with a horizontal puller (Model P97; Sutter Instrument), and filled with a solution containing 0.125% Neurobiotin and (in mM): 140 caesium gluconate, 10 HEPES, 2 $MgCl_2$, 3 Na_2ATP, 0.3 GTP, 0.1 spermine, 1 QX-314 (pH 7.3, 280–285 m-osM). All chemicals and drugs were purchased from Sigma-Aldrich.

Nucleus accumbens medium spiny neurons from the core region were identified under visual guidance by using infrared-differential interference contrast (IR-DIC) videomicroscopy with a 40× water-immersion objective (BX51-WI; Olympus). The image was detected with an infrared-sensitive charge-coupled device camera and displayed on a monitor. Whole-cell patch-clamp recordings were performed with a computer-controlled amplifier (MultiClamp 700B; Axon Instruments), digitized (Digidata 1440; Axon Instruments) and acquired with Axoscope 10.1 (Axon Instruments) at a sampling rate of 10 kHz. The liquid junction potential was not corrected and electrode potentials were adjusted to zero before obtaining the whole-cell configuration. Nucleus accumbens medium spiny neuron synaptic responses were elicited by local electrical stimulation (0.05–0.30 mA square pulses of 0.3 ms duration delivered every 20 s) of excitatory inputs by using a bipolar electrode made from a pair of twisted Teflon-coated nichrome wires (tips separated by approximately 200 μm) and placed about 300 μm lateral to the recorded neurons. The intensity of stimulation was chosen from the minimum amount of current necessary to elicit a synaptic response with less than 15% variability in amplitude during baseline recording[49]. Only neurons that retained such synaptic response reliability during the subsequent 20 min of baseline recording were included in the present study. If the current intensity required was more than 0.3 mA, the neuron was discarded.

All recordings were conducted in voltage-clamp configuration at 33–35 °C in the absence of tetrodotoxin. Control and drug-containing aCSF were continuously oxygenated throughout the experiments. After 20–30 min of baseline recording, a solution containing the GluR2-lacking AMPA receptor antagonist Naspm (100–200 μM) was perfused for 10 min, followed by a 20–30-min washout period. Changes in input resistance, spontaneous EPSC (frequency and amplitude), evoked EPSC amplitude and paired-pulse ratio (at 50 ms interval) were analysed before and after drug application. In addition, we collected several points of the current-voltage relationship (holding V_m at −70, −50, −30, +20, +40 and +60 mV) of the evoked AMPA-mediated EPSC during baseline to compute the rectification index. The rectification index (i_r) was calculated by correcting any potential shifts in the reversal potential values (E_{rev}) (ref. 50) and computed from the following equation: $i_r = [I_{-70}/(70 - E_{rev})]/[I_{+40}/(40 - E_{rev})]$. Thus, i_r is expressed as a ratio that increases when rectification increases. I_{-70} and I_{+40} are the EPSC current amplitudes recorded by holding the membrane potential at −70 mV and +40 mV, respectively. The E_{rev} values were obtained from the *I–V* relationship. Finally, we performed frequency and amplitude analyses of spontaneous AMPA receptor-mediated events by using Clampfit 10 (Axon Instruments). All comparisons were performed from 3-min segments of baseline recordings acquired at 10 kHz. For each neuron we assessed cumulative histograms and conducted Kolmogorov–Smirnov tests. All measures are expressed as means ± s.e.m. All neurons included in the present study were labelled with Neurobiotin. Their location and morphology were further confirmed as medium spiny neurons in the core region of the nucleus accumbens.

Statistical analyses. Data from self-administration experiments were analysed with the statistical program SPSS (GLM procedure). The nose-poke or lever-press data from the extinction tests for cue-induced cocaine-seeking were analysed by analysis of variance (ANOVA) with withdrawal day (1, 45) as the between-subjects factor, and hole or lever (previously active, inactive) as the within-subject factor. For the Naspm accumbens injection experiment, the statistical analyses also included the within-subjects factor of session hour. For biochemical studies, group differences in protein levels were analysed by ANOVA with drug exposure (saline, cocaine) or extinction test (yes, no) and withdrawal day (1, 45) as the between-subjects factors, followed by a post-hoc Tukey test. For experiments on the effect of Naspm on cocaine or sucrose self-administration, the ANOVA included the within-subjects factors of Naspm dose (vehicle, 40 μg) and session hour (1–6). For electrophysiological studies, drug effects were compared by using Student's *t*-test or repeated-measures ANOVA. Differences between experimental conditions were considered statistically significant when $P < 0.05$.

31. Paxinos, G. & Watson, C. *The Rat Brain in Stereotaxic Coordinates* (Elsevier Academic, Amsterdam, 2005).
32. Bossert, J. M., Poles, G., Wihbey, K., Koya, U. & Shaham, Y. Differential effects of blockade of dopamine D1-family receptors in nucleus accumbens core or shell on reinstatement of heroin seeking induced by contextual and discrete cues. *J. Neurosci.* **27**, 12655–12663 (2007).
33. Lu, L., Uejima, J. L., Gray, S. M., Bossert, J. M. & Shaham, Y. Systemic and central amygdala injections of the $mGluR_{2/3}$ agonist LY379268 attenuate the expression of incubation of cocaine craving. *Biol. Psychiatry* **61**, 591–598 (2007).
34. Takazawa, A. *et al.* Potent and long-lasting anticonvulsant effects of 1-naphthylacetyl spermine, an analogue of Joro spider toxin, against amygdaloid kindled seizures in rats. *Brain Res.* **706**, 173–176 (1996).
35. Noh, K. M. *et al.* Blockade of calcium-permeable AMPA receptors protects hippocampal neurons against global ischemia-induced death. *Proc. Natl Acad. Sci. USA* **102**, 12230–12235 (2005).
36. Hall, R. A., Hansen, A., Andersen, P. H. & Soderling, T. R. Surface expression of the AMPA receptor subunits GluR1, GluR2, and GluR4 in stably transfected baby hamster kidney cells. *J. Neurochem.* **68**, 625–630 (1997).
37. Hall, R. A. & Soderling, T. R. Differential surface expression and phosphorylation of the *N*-methyl-D-aspartate receptor subunits NR1 and NR2 in cultured hippocampal neurons. *J. Biol. Chem.* **272**, 4135–4140 (1997).
38. Hall, R. A. & Soderling, T. R. Quantitation of AMPA receptor surface expression in cultured hippocampal neurons. *Neuroscience* **78**, 361–371 (1997).
39. Archibald, K., Perry, M. J., Molnar, E. & Henley, J. M. Surface expression and metabolic half-life of AMPA receptors in cultured rat cerebellar granule cells. *Neuropharmacology* **37**, 1345–1353 (1998).
40. Broutman, G. & Baudry, M. Involvement of the secretory pathway for AMPA receptors in NMDA-induced potentiation in hippocampus. *J. Neurosci.* **21**, 27–34 (2001).
41. Clayton, D. A., Grosshans, D. R. & Browning, M. D. Aging and surface expression of hippocampal NMDA receptors. *J. Biol. Chem.* **277**, 14367–14369 (2002).

42. Grosshans, D. R., Clayton, D. A., Coultrap, S. J. & Browning, M. D. LTP leads to rapid surface expression of NMDA but not AMPA receptors in adult rat CA1. *Nature Neurosci.* **5,** 27–33 (2002).
43. Grosshans, D. R., Clayton, D. A., Coultrap, S. J. & Browning, M. D. Analysis of glutamate receptor surface expression in acute hippocampal slices. *Sci. STKE* **137,** PL8 (2002).
44. Gerges, N. Z. *et al.* Independent functions of hsp90 in neurotransmitter release and in the continuous synaptic cycling of AMPA receptors. *J. Neurosci.* **24,** 4758–4766 (2004).
45. Sans, N. *et al.* Aberrant formation of glutamate receptor complexes in hippocampal neurons of mice lacking the GluR2 AMPA receptor subunit. *J. Neurosci.* **23,** 9367–9373 (2003).
46. Wenthold, R. J., Petralia, R. S., Blahos, J. II & Niedzielski, A. S. Evidence for multiple AMPA receptor complexes in hippocampal CA1/CA2 neurons. *J. Neurosci.* **16,** 1982–1989 (1996).
47. Bernard, V., Somogyi, P. & Bolam, J. P. Cellular, subcellular, and subsynaptic distribution of AMPA-type glutamate receptor subunits in the neostriatum of the rat. *J. Neurosci.* **17,** 819–833 (1997).
48. Stefani, A. *et al.* Physiological and molecular properties of AMPA/kainate receptors expressed by striatal medium spiny neurons. *Dev. Neurosci.* **20,** 242–252 (1998).
49. Tseng, K. Y. & O'Donnell, P. D2 dopamine receptors recruit a GABA component for their attenuation of excitatory synaptic transmission in the adult rat prefrontal cortex. *Synapse* **61,** 843–850 (2007).
50. Balland, B., Lachamp, P., Strube, C., Kessler, J. P. & Tell, F. Glutamatergic synapses in the rat nucleus tractus solitarii develop by direct insertion of calcium-impermeable AMPA receptors and without activation of NMDA receptors. *J. Physiol. (Lond.)* **574,** 245–261 (2006).

LETTERS

Template-directed synthesis of a genetic polymer in a model protocell

Sheref S. Mansy[1], Jason P. Schrum[1], Mathangi Krishnamurthy[1], Sylvia Tobé[1], Douglas A. Treco[1] & Jack W. Szostak[1]

Contemporary phospholipid-based cell membranes are formidable barriers to the uptake of polar and charged molecules ranging from metal ions to complex nutrients. Modern cells therefore require sophisticated protein channels and pumps to mediate the exchange of molecules with their environment. The strong barrier function of membranes has made it difficult to understand the origin of cellular life and has been thought to preclude a heterotrophic lifestyle for primitive cells. Although nucleotides can cross dimyristoyl phosphatidylcholine membranes through defects formed at the gel-to-liquid transition temperature[1,2], phospholipid membranes lack the dynamic properties required for membrane growth. Fatty acids and their corresponding alcohols and glycerol monoesters are attractive candidates for the components of protocell membranes because they are simple amphiphiles that form bilayer membrane vesicles[3–5] that retain encapsulated oligonucleotides[3,6] and are capable of growth and division[7–9]. Here we show that such membranes allow the passage of charged molecules such as nucleotides, so that activated nucleotides added to the outside of a model protocell spontaneously cross the membrane and take part in efficient template copying in the protocell interior. The permeability properties of prebiotically plausible membranes suggest that primitive protocells could have acquired complex nutrients from their environment in the absence of any macromolecular transport machinery; that is, they could have been obligate heterotrophs.

Previous observations of the slow permeation of UMP across fatty-acid-based membranes[6] stimulated us to explore the structural factors that control the permeability of these membranes (Fig. 1). We examined membrane compositions with varied surface charge density, fluidity, and stability of regions of high local curvature. We began by studying the permeability of ribose, because this sugar is a key building block of the nucleic acid RNA, and because sugar permeability is conveniently measured with a real-time fluorescence readout of vesicle volume after solute addition[10,11]. We used pure myristoleic acid (C14:1 fatty acid, myristoleate in its ionized form) as a reference composition, because this compound generates robust vesicles that are more permeable to solutes than the more common longer chain oleic acid. Both myristoleyl alcohol and the glycerol monoester of myristoleic acid (monomyristolein, GMM) stabilize myristoleate vesicles to the disruptive effects of divalent cations[3,6]. Addition of these amphiphiles should decrease the surface charge density of myristoleate vesicles, whereas myristoleyl phosphate should increase the surface charge density. Surprisingly, only the addition of GMM affected ribose permeability, leading to a fourfold increase (Fig. 2a). This result suggested that surface charge density *per se* was not a major factor controlling sugar permeability.

We hypothesized that the larger steric bulk of the glycerol-ester head group of GMM relative to the carboxylate of myristoleic acid might increase ribose permeability by stabilizing highly curved surfaces associated with the formation of transient solute–lipid complexes[12]. We therefore examined the effect of the glycerol esters of the longer chain amphiphiles palmitoleic acid (C16:1) and oleic acid (C18:1) on the permeability of pure palmitoleic acid and oleic acid membranes. These molecules, which are progressively less cone-shaped than GMM, had an increasingly smaller influence on the permeability of the corresponding pure fatty acid membranes (Fig. 2b). However, the addition of sorbitan monooleate, which has a larger cyclic 6-carbon sugar head group (thus restoring a more conical shape to this 18-carbon fatty acid), resulted in a fourfold increase in the permeability of oleic acid membranes, consistent with the hypothesis that cone-shaped amphiphiles stabilize highly curved membrane deformations that facilitate solute passage. Decreasing acyl chain length within a series of homologous fatty acids (or mixtures of fatty acids and their glycerol esters) also led to increased sugar

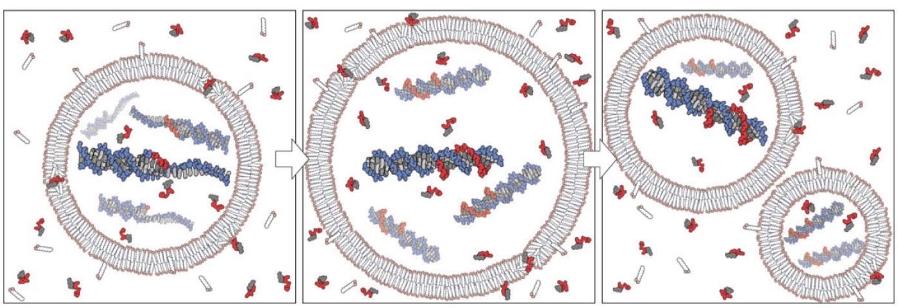


Figure 1 | Conceptual model of a heterotrophic protocell. Growth of the protocell membrane results from the incorporation of environmentally supplied amphiphiles, whereas division may be driven by intrinsic or extrinsic physical forces. Externally supplied activated nucleotides permeate across the protocell membrane and act as substrates for the non-enzymatic copying of internal templates. Complete template replication followed by random segregation of the replicated genetic material leads to the formation of daughter protocells.

[1]Howard Hughes Medical Institute, Department of Molecular Biology and the Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts 02114, USA.

permeability (Fig. 2b and Supplementary Table 1), presumably owing to the decreased stability of the ideal bilayer structure with respect to the formation of transient solute–lipid complexes.

To investigate further the idea that local membrane deformations are required for solute passage across the membrane, we asked whether increased packing disorder within the lipid bilayer would enhance permeability. Phospholipids with higher degrees of unsaturation yield more disordered, fluid membranes that are more permeable to water and small solutes[13]. We observed a fivefold increase in ribose permeability for vesicles composed of linoleic acid (C18:2) versus oleic acid (C18:1). Branched-chain amphiphiles such as the isoprenoid farnesol also increase the fluidity of phospholipid membranes[14]. Vesicles made from a 2:1 molar mixture of myristoleic acid and farnesol exhibited a ~17-fold increase in ribose permeability relative to pure myristoleic acid vesicles. Conversely, the higher packing density of saturated amphiphiles[13] should lead to increased membrane order and decreased solute permeability. As expected, the addition of lauric acid (C12:0) to myristoleic acid vesicles (2:1 myristoleic acid:lauric acid) resulted in a twofold decrease in ribose permeability (Fig. 2c).

The above experiments show that solute permeability can be increased by decreasing acyl chain length, by increasing acyl chain unsaturation or branching, and by adding amphiphiles with larger head groups. The most prebiotically plausible amphiphiles are the short chain saturated fatty acids and their corresponding alcohols and glycerol esters[15–17]. To see whether shorter chain length could compensate for the loss of unsaturation, we tested membrane compositions based on the C10 amphiphiles decanoic acid and decanol (DOH) as well as the glycerol monoester of decanoate (GMD). Pure decanoic acid only forms stable vesicles at very high amphiphile concentrations (≥100 mM), but the addition of decanol decreases the critical aggregate concentration to ~20 mM and increases the pH range over which vesicles are stable[3]. We found that the ribose permeability of 2:1 decanoate:decanol vesicles is very similar to that of myristoleic acid vesicles but is significantly less than that of myristoleic acid:GMM vesicles (Fig. 2d). Based on the observations that amphiphiles with larger head groups lead to increased permeability, we replaced half of the decanol with glycerol mono-decanoate. The resulting vesicles exhibited a tenfold increase in ribose permeability (Fig. 2d). It is particularly notable that improved permeability and stability are obtained with mixtures of amphiphiles, such as might be expected to be present in a chemically rich prebiotic environment. This is in marked contrast to the situation with nucleic acids, where homogeneous nucleotides are thought to be required for replication.

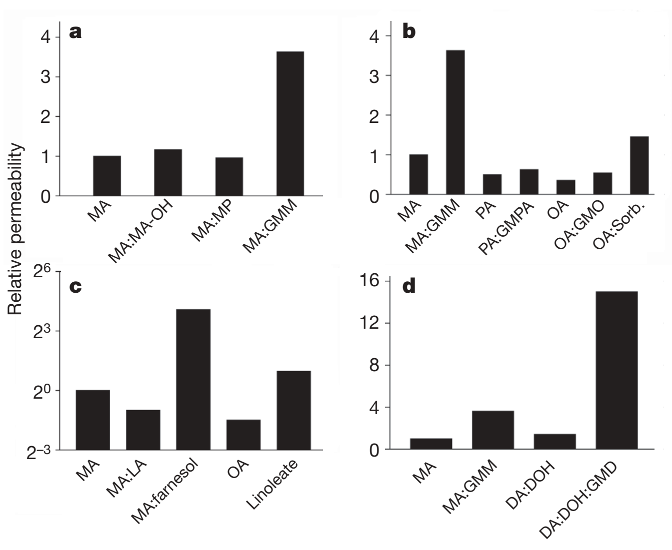


Figure 2 | Ribose permeability of fatty acid based membranes. **a–c**, Influence of head group charge (**a**), head group size (**b**) and membrane fluidity (**c**). **d**, Comparison of decanoic-acid-based membranes with myristoleic acid based membranes. All binary lipid mixtures were 2:1 molar ratios of fatty acid:additive; a 4:1:1 ratio of decanoic acid:DOH:GMD was used. Ribose permeabilities are relative to that of myristoleic acid membranes. DA, decanoic acid; DOH, decanol; GMD, glycerol monoester of decanoic acid; GMM, glycerol monoester of myristoleate; GMO, glycerol monoester of oleate; GMPA, glycerol monoester of palmitoleate; LA, lauric acid; MA, myristoleic acid; MA-OH, myristoleoyl alcohol; MP, myristoleoyl phosphate; OA, oleate; PA, palmitoleic acid; Sorb., sorbitan monooleate.

Vesicles with all of the above membrane compositions retained 100% of an encapsulated fluorescein-labelled dA_{10} oligonucleotide indefinitely (Supplementary Fig. 6). In addition, all membrane compositions retained the previously observed 3–10-fold faster permeation of ribose compared to its diastereomers arabinose, lyxose and xylose (Supplementary Table 1). These observations show that our permeability measurements do not reflect leakage of encapsulated materials caused by vesicle rupture or the formation of large non-selective pores.

Having established that membranes composed of prebiotically plausible amphiphilic molecules have high permeabilities to simple sugars, we asked whether such membranes would allow the uptake of nucleotide nutrients by a simple model protocell. We measured nucleotide permeation by encapsulating nucleotides within vesicles and then determining the fraction of the encapsulated nucleotide that had leaked out of the vesicles at various times. Because charge has such a dominant effect in restricting solute permeation through membranes, we first examined the effect of nucleotide charge on permeation through myristoleic acid:GMM (2:1) membranes. We observed negligible leakage of AMP, ADP or ATP (with 2, 3 and 4 negative charges at pH 8.5, respectively) over 24 h in the absence of Mg^{2+}, suggesting that these molecules were either too large or too highly charged to cross the membrane. We did observe slow permeation of AMP and ADP in the presence of 3 mM Mg^{2+} (Fig. 3a), as expected from the formation of complexes of reduced net charge[18]. The impermeability of ATP argues against a role for NTPs in very early forms of cellular life dependent on externally synthesized activated nucleotides; instead, NTPs may be a later evolutionary adaptation that prevents the leakage of internally synthesized activated nucleotides[19].

The above results highlight the importance of reducing the net charge of nucleotides to enhance membrane permeability. Imidazole-activated nucleotides have been used as convenient models of prebiotic activated nucleotides in studies of both spontaneous and templated polymerization reactions[20–23]. In addition to their higher intrinsic chemical reactivity compared to NTPs, these activated nucleotides are less polar and bear only a single negative charge at a neutral to moderately alkaline pH. We therefore measured the permeabilities of a series of adenosine nucleotides and their corresponding phosphorimidazolides, using both myristoleic acid:GMM (2:1) and C10 membranes (4:1:1 decanoic acid:DOH:GMD; Fig. 3b–d). The half-time for equilibration of nucleoside phosphorimidazolides using 100-nm vesicles was approximately 12 h. The effects of membrane composition on the permeability of nucleoside phosphorimidazolides were essentially parallel to our results for sugar permeability—pure myristoleic acid vesicles were less permeable to nucleotides than myristoleic acid:GMM (2:1) vesicles, whereas farnesol led to an even greater enhancement of permeability (Supplementary Fig. 4). Similarly, the permeability of decanoic acid:DOH membranes was enhanced by the addition of GMD (Fig. 3d).

Our permeability data are consistent with a transport model in which polar functional groups of solute molecules initially interact with one or more amphiphile head groups with displacement of bound water molecules (Supplementary Fig. 5), whereas non-polar regions of the solute may interact with the hydrophobic acyl chains of the amphiphiles[24]. Formation of this relatively non-specific amphiphile–solute complex is followed by a concerted inversion of the complex across the membrane. Lipids with large head groups could increase solute permeation by providing more opportunity for solute

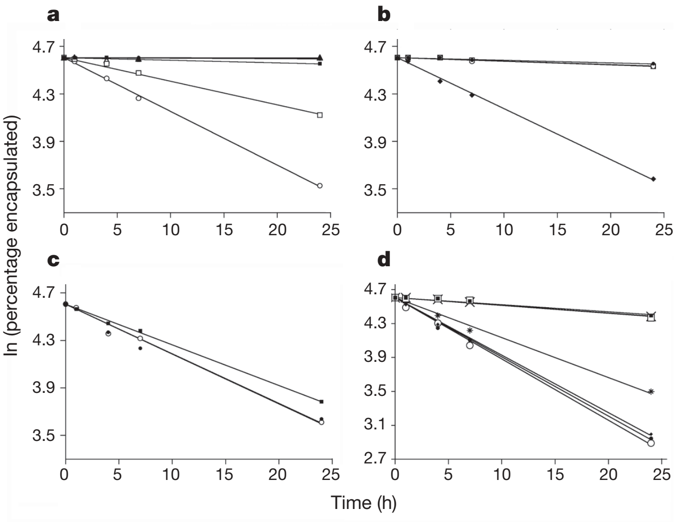


Figure 3 | Time courses of nucleotide permeation through fatty acid based membranes. a, Nucleotide permeation across MA:GMM membranes. Filled square, AMP; open square, AMP + 3 mM $MgCl_2$; filled circle, ADP; open circle, ADP + 3 mM $MgCl_2$, filled triangle, ATP; open triangle, ATP + 3 mM $MgCl_2$. **b**, Permeation of AMP derivatives across MA:GMM membranes. Filled circle, adenosine-5′-monophosphate; open circle, 2′-deoxyadenosine-5′-monophosphate; open square, 2′-amino-2′,3′-dideoxyadenosine-5′-monophosphate; filled diamond, 2′-deoxyadenosine-5′-phosphorimidazolide. **c**, Permeability of activated nucleotides across MA:GMM membranes. Filled square, adenosine-5′-phosphorimidazolide; filled circle, 2′-amino-2′,3′-dideoxyadenosine-5′-phosphorimidazolide; open circle, 3′-amino-3′-deoxyadenosine-5′-phosphorimidazolide. **d**, Nucleotide permeation across DA:DOH:GMD and DA:DOH membranes. Filled square, AMP; open square, dAMP; X, 2′-amino-2′,3′-dideoxyadenosine-5′-monophosphate; filled diamond and asterisk, adenosine-5′-phosphorimidazolide; filled circle, 2’-deoxyadenosine-5′-phosphorimidazolide; open circle, 2′-amino-2′,3′-dideoxyadenosine-5′-phosphorimidazolide. All are for 4:1:1 DA:DOH:GMD membranes except for the asterisk, which is 2:1 DA:DOH.

interaction, by favouring high local curvature and by decreasing the cohesive interactions between adjacent acyl chains and thereby facilitating amphiphile flip-flop. This model is similar to the previously proposed carrier model for the spontaneous transport of monovalent ions across fatty acid[25] and phospholipid membranes[26].

Encouraged by the observed permeability of activated nucleotides, we asked whether such nucleotides added to the outside of a model protocell could diffuse to the inside and engage in template-copying reactions in the vesicle interior. Although no sequence-general means for the non-enzymatic replication of a genetic polymer has yet been found, we have identified a system that exhibits remarkably rapid and efficient non-enzymatic copying of an oligo-dC DNA template (Supplementary Fig. 7). We used this system to model the spontaneous chemical replication of genetic material within a protocell. In brief, a DNA primer bearing a single 3′-amino-nucleotide at its 3′ terminus[27] was annealed to a DNA oligonucleotide consisting of a primer-binding region and a $(dC)_{15}$ template region. After the addition of 2′-amino, 2′-3′-dideoxyguanosine 5′-phosphorimidazolide, the primer was extended by the template-directed synthesis of 2′-phosphoramidate-linked DNA. Both 3′- and 2′-amino nucleotides polymerized much more rapidly than similarly activated ribo- or deoxyribo-nucleotides owing to the presence of the more nucleophilic amino group[23]. In solution, primer-extension across a $(dC)_{15}$ template in the presence of 5 mM activated 2′-amino-guanosine is essentially complete within 6 h (Fig. 4a). The principal product is full-length extended primer.

We used the reaction described previously to test the chemical and physical compatibility of template-directed copying with the integrity of fatty-acid-based vesicles. We examined the same template-copying reaction inside two sets of vesicles: the robust laboratory model system consisting of myristoleic acid:GMM (2:1) vesicles, and the more prebiotically plausible decanoic acid:DOH:GMD (4:1:1) vesicles. Vesicles containing encapsulated primer-template were purified to remove unencapsulated primer-template. We added 5 mM activated 2′-amino-guanosine to initiate template copying, removed aliquots at intervals, and again purified the vesicles to remove traces of primer-template that might have leaked out of the vesicles. The absence of measurable leakage of oligonucleotides from the vesicles shows that the activated nucleotides do not disrupt vesicle structure. Analysis of the reaction products showed significant primer-extension by 3 h, with full-length product continuing to accumulate until 24 h, at which point the vesicle reactions had reached a level of full-length product comparable to that seen in the solution reactions (Fig. 4). Thus, myristoleic acid:GMM or decanoic acid:DOH:GMD membranes slow the interaction between the primer-template and activated nucleotides, but are nevertheless compatible with template-copying chemistry in the vesicle interior. As expected, a similar experiment using myristoleic acid:farnesol (2:1) vesicles also showed efficient copying of encapsulated template (Fig. 4c). In contrast, phospholipid vesicles showed no detectable primer extension after the addition of activated nucleotide to the vesicle exterior (Fig. 4d).

The results described above bear directly on the two current contrasting views of the nature of protocells: the autotrophic and

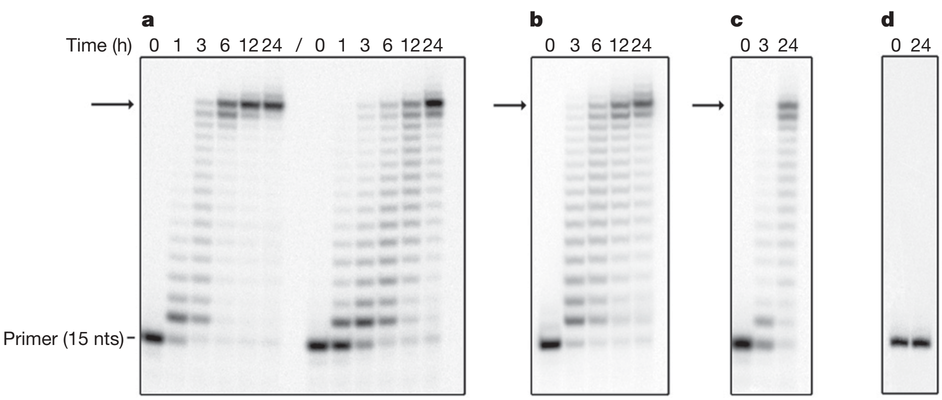


Figure 4 | Template-copying chemistry inside vesicles. Vesicles contained encapsulated primer-template complexes, and template-copying was initiated by the addition of activated monomer to the external solution. nts, nucleotides. **a**, Non-enzymatic dC_{15}-template copying in solution (lanes 1–6) and inside 2:1 MA:GMM vesicles (lanes 8–13) at 4 °C. **b**, Template-copying reaction in 4:1:1 DA:DOH:GMD vesicles at 25 °C. **c**, Template-copying reaction in 2:1 MA:farnesol vesicles at 4 °C. **d**, Template-copying reaction in POPC vesicles at 4 °C. For **a**–**c**, the arrow denotes full-length product. See Methods for reaction conditions.

heterotrophic models[28–30]. The autotrophic or 'metabolism first' model is based on the idea that autocatalytic reaction networks evolved in a spatially localized manner to generate *in situ* the building blocks required for cellular replication. Our results argue that early protocells with fatty-acid-based membranes could not have been autotrophs, because internally generated metabolites would leak out. In contrast, the heterotrophic model posits the emergence of very simple cellular structures within a complex environment that provides external sources of nutrients and energy. Although both models must overcome numerous conceptual difficulties related to the origin of complex molecular building blocks, the heterotrophic model was thought to face the additional difficulty of importing polar and even charged molecules across a bilayer lipid membrane. We have shown that fatty-acid-based membranes allow a simple protocell to acquire critical nutrients, while retaining polymerized nucleic acids indefinitely. Our results therefore support the idea that extremely simple heterotrophic protocells could have emerged within a prebiotic environment rich in complex nutrients.

METHODS SUMMARY

Sugar permeability. Vesicles were prepared with 10 mM encapsulated calcein in either 0.1 M POPSO (piperazine-1,4-bis(2-hydroxypropanesulfonic acid) dihydrate) and 3 mM EDTA (pH 8.2) or 0.1 M POPSO and 3 mM $MgCl_2$ (pH 8.2). Final sugar concentrations were either 0.5 M or 0.1 M. Permeability was measured by the shrink–swell assay[11] on an Applied Photophysics SX.18MV-R stopped-flow spectrometer at 23 °C.

Nucleotide permeability. Nucleotide permeability measurements were in 0.2 M sodium bicine (pH 8.5) at 23 °C and measured either by monitoring the leakage of entrapped nucleotide by radioactivity or by ultraviolet absorption. Separation of vesicle-entrapped and -released nucleotide was by gel filtration.

Primer extension reactions. Reactions contained 0.1 μM ^{32}P-labelled 3′-amino-terminated primer, 0.5 μM template DNA, 100 mM 1-(2-hydroxyethyl)-imidazole and 200 mM sodium bicine (pH 8.5). Reactions were initiated by the addition of 5 mM 2′-amino-2′,3′-dideoxyguanosine-5′-phosphorimidazolide and incubated at 4 °C. Samples were analysed by electrophoresis on a denaturing 17% polyacrylamide gel. Reaction products were visualized using a Typhoon 9410 PhosphorImager.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 29 October 2007; accepted 22 April 2008.
Published online 4 June 2008.

1. Chakrabarti, A. C., Breaker, R. R., Joyce, G. F. & Deamer, D. W. Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. *J. Mol. Evol.* **39,** 555–559 (1994).
2. Monnard, P. A., Luptak, A. & Deamer, D. W. Models of primitive cellular life: polymerases and templates in liposomes. *Phil. Trans. R. Soc. Lond. B* **362,** 1741–1750 (2007).
3. Apel, C. L., Deamer, D. W. & Mautner, M. N. Self-assembled vesicles of monocarboxylic acids and alcohols: conditions for stability and for the encapsulation of biopolymers. *Biochim. Biophys. Acta* **1559,** 1–9 (2002).
4. Blochliger, E., Blocher, M., Walde, P. & Luisi, P. L. Matrix effect in the size distribution of fatty acid vesicles. *J. Phys. Chem. B* **102,** 10383–10390 (1998).
5. Hargreaves, W. R. & Deamer, D. W. Liposomes from ionic, single-chain amphiphiles. *Biochemistry* **17,** 3759–3768 (1978).
6. Chen, I. A., Salehi-Ashtiani, K. & Szostak, J. W. RNA catalysis in model protocell vesicles. *J. Am. Chem. Soc.* **127,** 13213–13219 (2005).
7. Chen, I. A., Roberts, R. W. & Szostak, J. W. The emergence of competition between model protocells. *Science* **305,** 1474–1476 (2004).
8. Chen, I. A. & Szostak, J. W. A kinetic study of the growth of fatty acid vesicles. *Biophys. J.* **87,** 988–998 (2004).
9. Hanczyc, M. M., Fujikawa, S. M. & Szostak, J. W. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science* **302,** 618–622 (2003).
10. Chen, P. Y., Pearce, D. & Verkman, A. S. Membrane water and solute permeability determined quantitatively by self-quenching of an entrapped fluorophore. *Biochemistry* **27,** 5713–5718 (1988).
11. Sacerdote, M. G. & Szostak, J. W. Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proc. Natl Acad. Sci. USA* **102,** 6004–6008 (2005).
12. Israelachvili, J. N. *Intermolecular and Surface Forces* (Academic, London, 1992).
13. Lande, M. B., Donovan, J. M. & Zeidel, M. L. The relationship between membrane fluidity and permeabilities to water, solutes, ammonia, and protons. *J. Gen. Physiol.* **106,** 67–84 (1995).
14. Rowat, A. C., Keller, D. & Ipsen, J. H. Effects of farnesol on the physical properties of DMPC membranes. *Biochim. Biophys. Acta* **1713,** 29–39 (2005).
15. Deamer, D. W. Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* **317,** 792–794 (1985).
16. Huang, Y. *et al.* Molecular and compound-specific isotopic characterization of monocarboxylic acids in carbonaceous meteorites. *Geochim. Cosmochim. Acta* **69,** 1073–1084 (2005).
17. McCollom, T. M., Ritter, G. & Simoneit, B. R. Lipid synthesis under hydrothermal conditions by Fischer–Tropsch-type reactions. *Orig. Life Evol. Biosph.* **29,** 153–156 (1999).
18. Khalil, M. M. Complexation equilibria and determination of stability constants of binary and ternary complexes with ribonucleotides (AMP, ADP, and ATP) and salicylhydroxamic acid as ligands. *J. Chem. Eng. Data* **45,** 70–74 (2000).
19. Westheimer, F. H. Why nature chose phosphates. *Science* **235,** 1173–1178 (1987).
20. Eschenmoser, A. The search for the chemistry of life's orgin. *Tetrahedron* **63,** 12821–12844 (2007).
21. Ferris, J. P., Hill, A. R., Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381,** 59–61 (1996).
22. Kozlov, I. A., Pitsch, S. & Orgel, L. E. Oligomerization of activated D- and L-guanosine mononucleotides on templates containing D- and L-deoxycytidylate residues. *Proc. Natl Acad. Sci. USA* **95,** 13448–13452 (1998).
23. Tohidi, M., Zielinski, W. S., Chen, C. H. & Orgel, L. E. Oligomerization of the 3′-amino-3′deoxyguanosine-5′phosphorimidazolidate on a d(CpCpCpCpC) template. *J. Mol. Evol.* **25,** 97–99 (1987).
24. Wilson, M. A. & Pohorille, A. Mechanism of unassisted ion transport across membrane bilayers. *J. Am. Chem. Soc.* **118,** 6580–6587 (1996).
25. Chen, I. A. & Szostak, J. W. Membrane growth can generate a transmembrane pH gradient in fatty acid vesicles. *Proc. Natl Acad. Sci. USA* **101,** 7965–7970 (2004).
26. Paula, S. G., Volkov, A. G., Van Hoek, A. N., Haines, T. H. & Deamer, D. W. Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys. J.* **70,** 339–348 (1996).
27. Hagenbuch, P., Kervio, E., Hochgesand, A., Plutowski, U. & Clemens, R. Chemical primer extension: efficiently determining single nucleotides in DNA. *Angew. Chem. Int. Edn Engl.* **44,** 6588–6592 (2005).
28. Chen, I. A., Hanczyc, M. M., Sazani, P. L. & Szostak, J. W. in *The RNA World* (eds Gesteland, R. F., Cech, T. R. & Atkins, J. F.) 57–88 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 2006).
29. Morowitz, H. J. *Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis* (Yale Univ. Press, New Haven, 2004).
30. Wachtershauser, G. Evolution of the first metabolic cycles. *Proc. Natl Acad. Sci. USA* **87,** 200–204 (1990).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by grants from the NASA Exobiology Program (EXB02-0031-0018) and the NSF (CHE-0434507) to J.W.S. J.W.S. is an Investigator of the Howard Hughes Medical Institute. S.S.M. was supported by the NIH (F32 GM07450601). We thank I. Chen, M. Hanczyc, R. Bruckner, T. Zhu and Q. Dufton for discussions, and J. Iwasa for Fig. 1 and Supplementary Fig. 5.

Author Contributions Permeability experiments were performed by S.S.M. J.P.S. performed primer-extension experiments. M.K. synthesized 2′-aminoguanosine. S.T. and D.A.T. contributed to the development of the encapsulated primer-extension system. All authors helped to design the experiments and discussed the results. S.S.M., J.P.S. and J.W.S. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.W.S. (szostak@molbio.mgh.harvard.edu).

METHODS

Materials. Fatty acids, fatty alcohols and the glycerol monoesters of fatty acids were obtained from Nu-chek Prep, Inc. POPC (1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine) was from Avanti Polar Lipids, Inc. Myristoleoyl phosphate[31–33] was synthesized as described previously. 2′-amino-2′,3′-dideoxyguanosine-5′-phosphorimidazolide was synthesized by first generating 2′-azido-2′,3′-dideoxyguanosine, as described previously[34], followed by phosphorylation of the 2′-azido-2′,3′-dideoxy nucleoside with $POCl_3$ in triethyl phosphate, activation with 1,1′-carbonyl-diimidazole (CDI) to yield the 5′-phosphorimidazolide, and reduction of the 2′-azido group to the 2′-amine by catalytic hydrogenation. Nucleotide phosphorimidazolides were then purified by reverse-phase HPLC on an Alltima C18 column (Alltech) equilibrated with 0.1 M triethylammonium bicarbonate/2% acetonitrile, pH 8.0, and eluted with an acetonitrile gradient. Oligonucleotides were synthesized on an Expedite 8909 DNA synthesizer (Applied Biosystems). Template DNA (5′-AACCCCCCC CCCCCCCCCCAGTCAGTCTACGC-3′) for primer extension reactions was synthesized using standard phosphoramidite chemistry. 3′-amino-terminated DNA primer (5′-GCGTAGACTGACTGG-NH_2-3′) was synthesized using reverse phosphoramidites (Glen Research), with the final addition using a 3′-amino phosphoramidite (Transgenomic). Oligonucleotides were purified by anion exchange HPLC on a DNAPac PA-100 column (Dionex) in 0.01 M NaOH/0.01 M NaCl, pH 12.0, in a gradient up to 1.5 M NaCl.

Vesicle preparation. Fatty acid vesicles were prepared by oil dispersion in buffered solutions as described previously[1,3]. For vesicles composed of mixtures of unsaturated amphiphiles, the oils were mixed before dispersion in aqueous solution. Vesicles of mixed saturated and unsaturated composition were made by first generating vesicles composed of the unsaturated amphiphile, extrusion through 100 nm pore-size polycarbonate filters, followed by the addition of micelles composed of the saturated fatty acid. All vesicle preparations were extruded 11 times with an Avanti mini-extruder. For the encapsulation of molecules, amphiphiles were resuspended in the presence of the encapsulant followed by freeze-thaw cycling to equilibrate internal and external solutes. Separation of entrapped and unencapsulated material was by gel filtration with Sepharose-4B resin (Sigma-Aldrich) in which the running buffer contained the same amphiphile composition as the vesicles at a concentration above their critical aggregate concentration. Vesicle size was measured by dynamic light scattering with a PDDLS/CoolBatch 90T from Precision Detectors.

Sugar permeability. Vesicles were prepared with 10 mM encapsulated calcein in either 0.1 M POPSO and 3 mM EDTA, pH 8.2, or 0.1 M POPSO and 3 mM $MgCl_2$, pH 8.2. Final sugar concentrations were either 0.5 M or 0.1 M. Before measurement, vesicle samples were diluted to 4 mM amphiphile in buffer containing amphiphiles of equivalent composition as the vesicle above its critical aggregate concentration (myristoleic-acid-containing vesicles, 4 mM; palmitoleic-acid-containing vesicles, 1 mM; oleic-acid-containing vesicles, 0.1 mM; decanoic-acid-containing vesicles, 20 mM). Permeability was measured by the shrink–swell assay[13] on an Applied Photophysics SX.18MV-R stopped-flow spectrometer at 23 °C. The rate of the initial volume decrease due to water efflux yields the water permeability P_w, and the rate of the slower relaxation back to the initial volume reflects solute entry and yields the solute permeability P_s. Excitation and emission were at 470 nm and 540–560 nm, respectively. To avoid inner-filter effects and interferences arising from scattered light, all samples had absorbance values at 470 nm and 600 nm < 0.1. Size-exclusion chromatography showed that no calcein leaked out of the vesicles during the stopped-flow experiments.

Nucleotide permeability. Nucleotide permeability measurements were in 0.2 M sodium bicine, pH 8.5, at 23 °C and were measured either by monitoring the leakage of entrapped nucleotide by radioactivity or by ultraviolet absorption. The leakage of radioactive nucleotide was measured by loading aliquots at different time points on a gel filtration column and analysing fractions by scintillation counting. Permeability measurements of non-radioactive nucleotides were similarly performed, except that quantification relied on 260 nm absorbance following twofold dilution of the fractions with methanol.

Primer extension reactions. Reactions contained 0.1 μM ^{32}P-labelled 3′-amino-terminated primer, 0.5 μM template DNA, 100 mM 1-(2-hydroxyethyl)-imidazole, and 200 mM sodium bicine, pH 8.5. Reactions were initiated by the addition of 5 mM 2′-amino-2′,3′-dideoxyguanosine-5′-phosphorimidazolide and incubated at 4 °C. Solution reactions were stopped by adding three volumes formamide and heating to 95 °C for 10 min followed by ethanol precipitation. Vesicle reactions were stopped by gel filtration followed immediately by the addition of 0.3% Triton X-100 and ethanol precipitation. Stopped reactions were then resuspended in formamide gel loading buffer and heated to 95 °C for 2 min. Samples were analysed by electrophoresis on a denaturing 17% polyacrylamide gel. Reaction products were visualized using a Typhoon 9410 PhosphorImager. 1-(2-hydroxyethyl)imidazole enhances both non-enzymatic polymerization and nucleotide permeability by about twofold without affecting membrane integrity (Supplementary Figs 2 and 3). We confirmed that the primer was extended with phosphoramidate-linked G residues by the expected sensitivity to acid hydrolysis; in separate experiments with a shorter primer and template, we confirmed the non-enzymatic synthesis of phosphoramidate linked DNA by matrix assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF-MS).

Vesicle stability. The stability of vesicles of different compositions was assessed by quantifying leakage of entrapped 5′-fluorescein-labelled dA_{10} (Massachusetts General Hospital DNA Core Facility) after 24 h at 23 °C in 0.2 M sodium bicine, pH 8.5. Vesicles were separated from leaked oligonucleotides by gel filtration chromatography (Sepharose 4B) and quantified by fluorescence ($\lambda^{excitation} = 490$ nm, $\lambda^{emission} = 520$ nm) with a SpectraMAX GeminiEM fluorescence plate reader (Molecular Devices). To test the influence of 1-(2-hydroxyethyl)imidazole on vesicle stability, 2:1 myristoleic acid:GMM vesicle solutions were supplemented with 100 mM 1-(2-hydroxyethyl)imidazole and tested as described previously.

31. Danilov, L. L. & Chojnacki, T. A simple procedure for preparing dolichyl monophosphate by the use of $POCl_3$. *FEBS Lett.* **131,** 310–312 (1981).
32. Guernelli, S. *et al.* Supramolecular complex formation: a study of the interactions between β-cyclodextrin and some different classes of organic compounds by ESI-MS, surface tension measurements, and UV/Vis and ^{1}H NMR spectroscopy. *Eur. J. Org. Chem.* **24,** 4765–4776 (2003).
33. Nelson, A. K. & Toy, A. D. F. The preparation of long-chain monoalkyl phosphates from pyrophosphoric acid and alcohols. *Inorg. Chem.* **2,** 775–777 (1963).
34. Kawana, M. & Kuzuhara, H. General method for the synthesis of 2′-azido-2′,3′-dideoxynucleosides by the use of [1,2]-hydride shift and β-elimination reactions. *J. Chem. Soc. Perkin Trans. I* **4,** 469–478 (1992).

Vol 454 | 3 July 2008 | doi:10.1038/nature06992

LETTERS

Induced ncRNAs allosterically modify RNA-binding proteins *in cis* to inhibit transcription

Xiangting Wang[1,2], Shigeki Arai[5]*, Xiaoyuan Song[1]*, Donna Reichart[3], Kun Du[5], Gabriel Pascual[3,4], Paul Tempst[6], Michael G. Rosenfeld[1,4], Christopher K. Glass[3,4] & Riki Kurokawa[5]

With the recent recognition of non-coding RNAs (ncRNAs) flanking many genes[1–5], a central issue is to obtain a full understanding of their potential roles in regulated gene transcription programmes, possibly through different mechanisms[6–12]. Here we show that an RNA-binding protein, TLS (for translocated in liposarcoma), serves as a key transcriptional regulatory sensor of DNA damage signals that, on the basis of its allosteric modulation by RNA, specifically binds to and inhibits CREB-binding protein (CBP) and p300 histone acetyltransferase activities on a repressed gene target, *cyclin D1* (*CCND1*) in human cell lines. Recruitment of TLS to the *CCND1* promoter to cause gene-specific repression is directed by single-stranded, low-copy-number ncRNA transcripts tethered to the 5′ regulatory regions of *CCND1* that are induced in response to DNA damage signals. Our data suggest that signal-induced ncRNAs localized to regulatory regions of transcription units can act cooperatively as selective ligands, recruiting and modulating the activities of distinct classes of RNA-binding co-regulators in response to specific signals, providing an unexpected ncRNA/RNA-binding protein-based strategy to integrate transcriptional programmes.

Transcriptional co-regulators, including coactivators and co-repressors, are required for the regulation of programmes of gene expression in a transcription factor-specific and gene-specific manner[13,14]. Among them, the histone acetyltransferases (HATs) CBP and p300 are essential as coactivators of multiple classes of signal-dependent transcription factors[13,14]. To search for cellular factors that might regulate the HAT activity of CBP, we incubated HeLa whole-cell extracts with full-length, Flag-tagged CBP immobilized on anti-Flag IgG affinity beads (Supplementary Fig. 1a) and observed a marked inhibition of CBP HAT activity on histones (Fig. 1a). Subcellular fractionation studies indicated the presence of two classes of inhibitory activity: one that bound to CBP and was present primarily in nuclear extracts (Fig. 1a, lane 3), and the other, the INHAT complex[15], that was present in both nuclear and cytoplasmic extracts (Supplementary Fig. 2a).

The nuclear activity that inhibited CBP in pull-down HAT assays fractionated as two main peaks by gel-filtration chromatography (Fig. 1b, top, and Supplementary Fig. 1b). Pooled fractions were further purified with full-length, Flag-tagged CBP linked to anti-Flag IgG beads, on the basis of the observation that inhibitory activity was observed with full-length CBP but not with the isolated HAT domain (Supplementary Fig. 2b). A large number of proteins were recovered from the high-molecular-mass fractions and a main band of about 75 kDa in the low-molecular-mass fractions (Fig. 1c). with matrix-assisted laser desorption/ionization reflectron time-of-flight mass spectrometric (MALDI-re-TOF MS) analysis[16], this 75-kDa protein was identified in three independent purifications as TLS, an RNA-binding protein that has been suggested to function in transcription[17], RNA processing[18] and DNA repair[19–22].

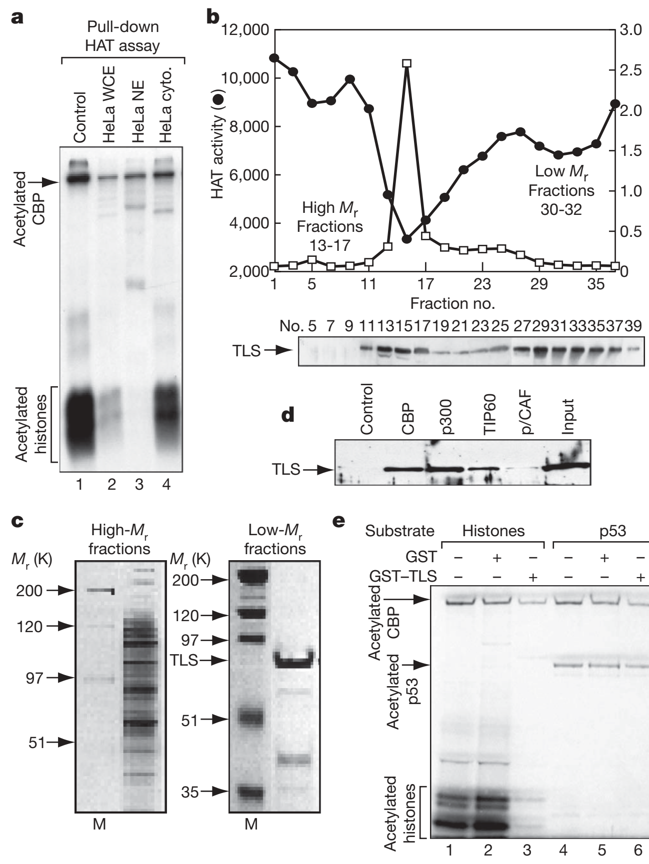


Figure 1 | TLS is a specific inhibitor of CBP and p300 HAT activity. a, CBP HAT activity measured by pull-down HAT assay. WCE, whole-cell extract; NE, nuclear extract; cyto., cytoplasmic extract. **b**, Top: inhibitory activity towards CBP HAT revealed by gel-filtration chromatography. M_r, molecular mass. Bottom: profile of TLS detected by western blotting. **c**, Representative silver-stained gels of pooled high-molecular-mass and low-molecular-mass fractions. M, molecular mass standards. **d**, TLS interacts with CBP, p300 and TIP60 but not with p/CAF. **e**, The effect of CBP HAT activity on histones or p53, determined by using GST–TLS.

[1]Howard Hughes Medical Institute, [2]Molecular Pathology Graduate Program, [3]Department of Cellular and Molecular Medicine, and [4]Department of Medicine, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA. [5]Division of Gene Structure and Function, Research Center for Genomic Medicine, Saitama Medical University, 1397-1 Yamane, Hidaka-shi, Saitama-Ken, Mail code 350-1241, Japan. [6]Molecular Biology Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA.
*These authors contributed equally to this work.

These findings were extended by demonstrating that recombinant TLS could bind to CBP (Fig. 1d) and strongly inhibited CBP HAT activity on core histones (Fig. 1e, lane 3). Glutathione *S*-transferase (GST)–TLS partly inhibited the acetylation of CBP itself, but not that of p53 (Fig. 1e, lane 6), suggesting that TLS selectively inhibits the ability of the acetylated CBP to transfer acetate to specific substrates. TLS also bound to p300 and TIP60 (for Tat-interacting protein 60) with similar affinities, but not to p/CAF (for p300/CBP-associated factor) (Fig. 1d and Supplementary Fig. 2c). GST–TLS inhibited the HAT activity of p300 (Fig. 2b) but not that of TIP60 (Supplementary Fig. 2d, e). TLS was also able to inhibit the acetylation by CBP of histones in nucleosomes prepared from HeLa cell nuclei (Supplementary Fig. 2f). TLS and its two related proteins EWS (for Ewing's sarcoma) and TAFII68 (for TATA-box-binding-protein-associated factor 68) (ref. 23) all proved to be present in high-molecular-mass fractions that correlate with activity that is inhibitory of CBP HAT (Fig. 1b, bottom, and Supplementary Fig. 3a). Similarly, EWS and TAFII68 were found to bind to CBP and TIP60 but not to p/CAF (Supplementary Fig. 3b, d), and exerted inhibitory effects on HAT activities of CBP and p300 (Supplementary Fig. 3c; data not shown). TLS interacted with several regions of CBP, with the region including the p160-interaction domain[24] (residues 1892–2441) serving as the most effective interaction domain (Supplementary Fig. 4). Pull-down HAT assays showed that recombinant TLS had no effect on the HAT activity of the isolated CBP_HAT region (Supplementary Fig. 2g), suggesting that the weak interaction of TLS with the CBP HAT domain (residues 1099–1877) is not sufficient for inhibitory effects towards HAT.

We next tested whether the inhibition of CBP HAT by TLS was RNA dependent. A synthetic RNA containing the consensus sequence GGUG (referred to below as the GGUG oligonucleotide) bound to TLS; mutations of GGUG to CCUC caused impaired binding[25] (Supplementary Fig. 5a). Treatment of TLS, EWS or TAFII68 with RNase A resulted in dissociation from p300 and CBP but not from TIP60 (Fig. 2a; Supplementary Fig. 5e, lanes 3–4; Supplementary Fig. 5f; Supplementary Fig. 6). Consistently, the inhibitory activity of GST–TLS on p300 HAT was abolished when GST–TLS was pretreated with the Ca^{2+}-dependent micrococcal nuclease (MNase) but not with DNase I (Fig. 2b). After blocking of MNase activity with EGTA, addition of the GGUG oligonucleotide, but not the CCUC oligonucleotide, restored the inhibitory effect of TLS on p300 HAT activity (Fig. 2b).

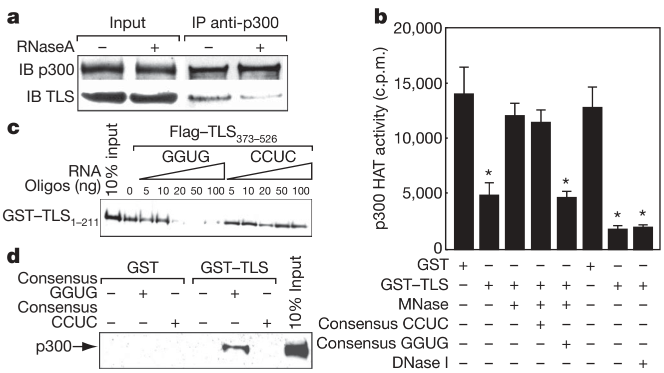


Figure 2 | Consensus GGUG-containing RNA oligonucleotide promotes the inhibitory effect of TLS on CBP and p300 HAT activities. **a**, Co-immunoprecipitation (IP) of p300 and TLS from HeLa cells treated with RNase A. **b**, p300 HAT activity was measured with micrococcal nuclease- (MNase) or DNase I-pretreated GST–TLS in the presence of GGUG or CCUC oligonucleotide. Asterisk, $P < 0.02$ compared with GST; $n = 3$. Error bars indicate s.e.m. **c**, **d**, Interaction between TLS N terminus (residues 1–211) and C terminus (residues 373–526) (**c**) or GST–TLS:p300 (**d**) in the presence of GGUG or CCUC oligonucleotide. GST and GST–TLS were pretreated with RNase A.

Interaction studies showed that the carboxy terminus of TLS (residues 211–526; 373–526 interacted with the GGUG oligonucleotide (Supplementary Fig. 5b), whereas the amino terminus (residues 1–211) interacted with CBP (Supplementary Fig. 5c).The N terminus of TLS was found to possess a detectably stronger inhibitory activity towards CBP HAT than did the full-length TLS (Supplementary Fig. 5d), and its interaction with CBP was not disrupted by treatment with RNase A (Supplementary Fig. 5e, lanes 1 and 2). Furthermore, the N terminus of TLS was capable of interacting with the C terminus of TLS (residues 373–526) in a manner that was inhibited by GGUG oligonucleotide in a dose-dependent manner (Fig. 2c), whereas the GGUG oligonucleotide enhanced the binding of TLS to p300 and CBP (Fig. 2d; data not shown). Partial proteolysis assays revealed that the GGUG oligonucleotide enhanced the cleavage of TLS (Supplementary Fig. 5g). Taken together, our findings suggest that an RNA-dependent allosteric modification of TLS relieves the inhibitory function of the C terminus of TLS, allowing the N terminus of TLS to bind to CBP and p300 and to regulate the HAT activity allosterically.

CCND1, a cell cycle regulator repressed by DNA damage signals[26], is an endogenous CREB target gene[27] and is induced in RAW264.7 cells by forskolin (Fig. 3a). Specific murine *TLS* short interfering RNA (siRNA; in this case referred to as siTLS) caused a marked increase in both basal and forskolin-stimulated *CCND1* mRNA levels in these cells (Fig. 3a and Supplementary Fig. 7a). Overexpression of human *TLS* could overcome the effect of siTLS (Supplementary Fig. 7c). Knockdown of p300 and CBP with specific siRNAs significantly decreased the acetylation of histone H3 on Lys 9 and Lys 14 (AcH3-K9K14) on the *CCND1* promoter detected by chromatin immunoprecipitation (ChIP) and *CCND1* mRNA levels (Fig. 3b, c, and

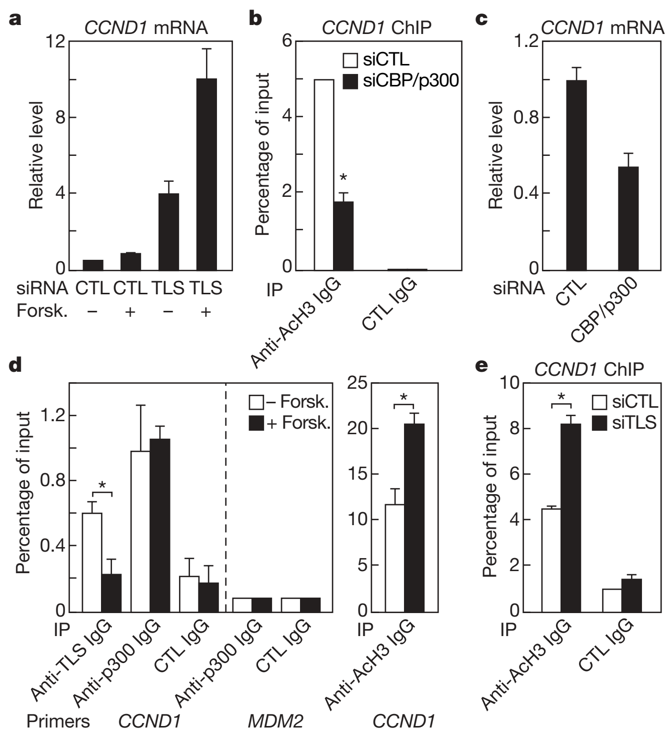


Figure 3 | TLS negatively regulates the CBP and p300 HAT-regulated *CCND1* gene. **a**, *CCND1* gene expression from RAW264.7 cells treated with forskolin (Forsk.) and *TLS* siRNA. CTL, control. **b**, **c**, Chromatin IP (ChIP) of histone acetylation (AcH3-K9K14) on the *CCND1* promoter (**b**) and *CCND1* gene expression (**c**) in the presence of control or *CBP* and *p300* siRNAs (siCBP/p300). Asterisk, $P < 0.01$; $n = 3$. **d**, ChIP with indicated immunoglobulin G (IgG) on the *CCND1* promoter on treatment with forskolin. *MDM2*, control. Asterisk, $P < 0.01$; $n = 3$. **e**, ChIP of AcH3-K9K14 on the *CCND1* promoter in the presence of control or TLS siRNA. Asterisk, $P < 0.01$; $n = 3$. Error bars indicate s.e.m.

Supplementary Fig. 7a), indicating required functions of these coactivators on this gene. Wild-type CBP, but not a HAT-mutant CBP[28], upregulated *CCND1* promoter activity (Supplementary Fig. 7b), suggesting that *CCND1* expression is dependent on the HAT function of CBP.

When RAW264.7 cells were cultured with carrier (without forskolin) in serum-starved medium, both p300 and TLS were bound to the *CCND1* promoter at the CRE site (Fig. 3d). Treatment with forskolin caused TLS to be dismissed from the *CCND1* promoter (Fig. 3d), despite a slight increase in total cellular levels of TLS (Supplementary Fig. 8a). In contrast, p300 remained bound (Fig. 3d). ChIP analysis revealed hyperacetylation of histone (AcH3-K9K14) on the *CCND1* promoter after treatment with forskolin (Fig. 3d) or knockdown of TLS (Fig. 3e). Taken together, our data suggest that TLS acts as a repressor of *CCND1*. However, we did not observe binding of TLS on all CREB targets (Supplementary Fig. 8b), suggesting that the negative regulation of CREB target genes by TLS is gene-specific.

In searching for endogenous regulatory RNAs, we took advantage of the fact that the expression of *CCND1* is downregulated in response to DNA damage signals such as those arising from ionizing radiation[29], correlated with decreased histone acetylation (Supplementary Fig. 9a, b). We considered previously unrecognized local transcripts, generated upstream of the *CCND1* promoter, as possible candidates. As shown in Fig. 4a, first-strand synthesis was performed with random primers, followed by real-time PCR with a series of validated specific primer pairs that showed similar amplification efficiencies on genomic DNA templates, spanning from −2008 to −162 base pairs upstream of the established *CCND1* transcription start site. These experiments revealed the presence of multiple previously unrecognized, ionizing-radiation-enhanced ncRNAs (A, B, D and E) transcribed from multiple 5′ regulatory regions of *CCND1* (ncRNA$_{CCND1}$; Fig. 4a). TLS interacted with these ncRNA$_{CCND1}$s as detected by RNA immunoprecipitation assays (Fig. 4b; data not shown), and a ChIP assay revealed that TLS was recruited to these ncRNA$_{CCND1}$-'expressing' regions in an ionizing-radiation-induced manner (Fig. 4c). In contrast, TLS showed very weak interaction with ncRNA$_{CCND1}$-'non-expressing' regions C and F (Fig. 4c). The level of TLS protein was never upregulated by ionizing irradiation, being either unchanged or, in some experiments, actually downregulated (Supplementary Fig. 9c).

Subcellular and chromatin fractionation studies revealed that ncRNA$_{CCND1}$ was mainly bound to chromatin (Fig. 4d). Real-time PCR analyses using several RNA species for which copy numbers have been well established as standards revealed that ncRNA$_{CCND1}$ was present at a low copy number (for example, region D at about two copies per cell under basal conditions and about four copies per cell after treatment with ionizing radiation; Fig. 4e). To test whether ncRNA$_{CCND1}$ might be present, in part, as an RNA–DNA hybrid, we evaluated the effects of treatment with RNase H and found that this treatment partly diminished ncRNA$_{CCND1}$ (Fig. 4f). A portion of ncRNAs was also diminished by RNase T1, which digests single-stranded RNA. The combination of RNase H and RNase T1 caused a complete loss of ncRNAs (Fig. 4f). This suggests that a portion of the ncRNA exists, at least transiently, as single-stranded RNA, in addition to a portion present as an RNA–DNA hybrid. TLS did not bind to the corresponding DNA sequence, nor did it bind to an RNA–DNA hybrid of the tested sequences (Fig. 4g; data not shown). ChIP for TLS on the *CCND1* promoter was performed after digestion with RNase H or RNase T1, or both. As shown in Fig. 4h, RNase T1 blocked TLS recruitment, whereas treatment with RNase H had no inhibitory effect. These data argue against RNA–DNA hybrids serving as the landing pads for TLS. Our data also revealed the presence of bidirectional ncRNA transcripts, further induced by ionizing radiation (Supplementary Fig. 10); in contrast, the adjacent 5′ untranslated region (UTR) of *CCND1* mRNA showed a decreased level in response to ionizing radiation (Supplementary Fig. 10).

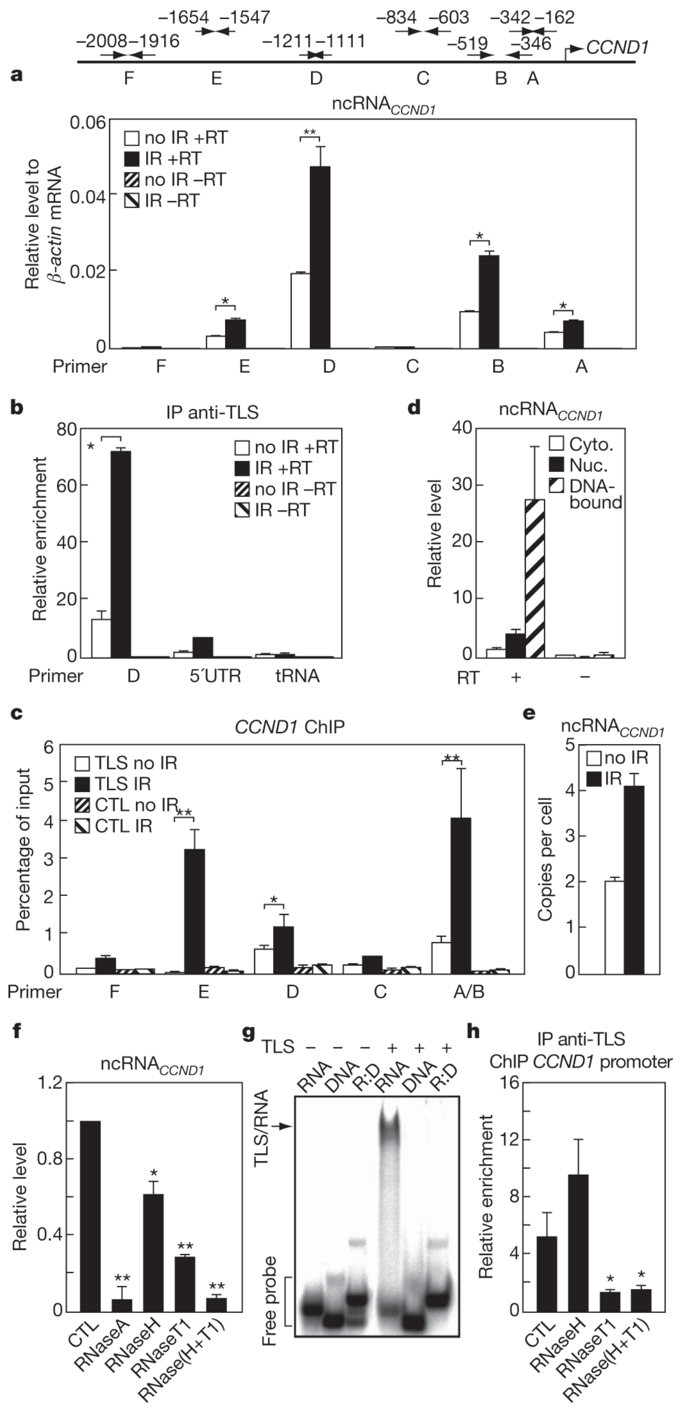


Figure 4 | ncRNA$_{CCND1}$s are predominantly single-stranded, DNA-bound species that bind to TLS. **a**, Top: diagram of ncRNA$_{CCND1}$ detection primers. Bottom: expression levels of ncRNA$_{CCND1}$s. IR, ionizing radiation; RT, reverse transcriptase. Asterisk, $P < 0.01$; two asterisks, $P < 0.002$; $n = 6$. **b**, Immunoprecipitation (IP) of TLS and detection of associated RNA by reverse transcription and real-time PCR. D, ncRNA$_{CCND1}$_D; 5′UTR, 5′ UTR of *CCND1*; transfer(t) RNA, tRNA14TyrATA. Asterisk, $P < 0.01$; $n = 3$. **c**, ChIP of TLS on the ncRNA$_{CCND1}$-'expressing' (E, D and AB) and ncRNA$_{CCND1}$-'non-expressing' regions (F and C). Asterisk, $P < 0.05$; two asterisks, $P < 0.01$; $n = 3$. CTL, control. **d**, Subcellular analysis of ncRNA$_{CCND1}$_D. **e**, Copy number of ncRNA$_{CCND1}$_D. **f**, Expression levels of ncRNA$_{CCND1}$_D on treatment with indicated RNases. Asterisk, $P < 0.05$; two asterisks, $P < 0.001$ compared with control, $n = 3$. **g**, Gel-shift analysis of TLS interactions with RNA, complementary DNA or RNA:DNA hybrid (R:D). RNA, −454s derived from ncRNA$_{CCND1}$_B. **h**, ChIP of TLS on the *CCND1* promoter on treatment with the indicated RNases. Asterisk, $P < 0.01$ compared with control; $n = 3$. Error bars indicate s.e.m.

Northern blotting analysis, with non-overlapped probes (about 200 nucleotides each) targeting the 5′ regulatory regions of *CCND1*, showed species of about 330 and about 200 nucleotides, and larger transcripts (Supplementary Fig. 11). The observations of clear variability in the lengths of these RNAs, and the fact that the bands were always multiple or diffuse, suggest diverse RNA polymerase II entry sites, or/and imprecise processing. ncRNA$_{CCND1}$ proved to be regulated by RNA polymerase II and polyadenylated, but not capped (Supplementary Fig. 12).

To investigate the potential function of ncRNA$_{CCND1}$, we identified specific siRNAs to the ncRNA$_{CCND1}$-'expressing' regions A (siA), D (siD) and E (siE); the ncRNA$_{CCND1}$-'non-expressing' regions C (siC) and F (siF); and the antisense 5′ UTR of *CCND1* (si5′UTR). SiA specifically knocked down ncRNA$_{CCND1}$ in region A without affecting that in region D; conversely, siD knocked down ncRNA$_{CCND1}$ in region D but not in region A (Supplementary Fig. 13), suggesting that multiple ncRNA transcripts were present, either as separate transcripts or as a result of rapid processing. Both strands of ncRNA$_{CCND1}$ were targeted by siA or siD (Supplementary Fig. 13). SiA, siD and siE (Fig. 5a, left), but not siC, siF or si5′UTR (Supplementary Fig. 15a), significantly enhanced the levels of endogenous *CCND1* mRNA. In contrast, the expression of *CCNE1* mRNA was not affected by either siA or siD (Fig. 5a, right). Cotransfection with siA, siD and siE (siADE) showed similar effects on the *CCND1* mRNA level compared with transfection with a single siRNA (Fig. 5a, left). As a control, the siRNA targeting the *CCND1* coding region (si*CCND1*) specifically blocked *CCND1* expression (Fig. 5a, left). These results argue against a *trans*-acting role for ncRNA$_{CCND1}$. SiD also enhanced the activity of *CCND1* promoter-driven reporter containing the ncRNA$_{CCND1}$-'expressing' region (Supplementary Fig. 14). In agreement with their putative local biological roles, siA or siD (Fig. 5b), but not siC or siF (Supplementary Fig. 15b), caused a decrease in TLS recruitment to the *CCND1* promoter at region A on treatment with ionizing radiation. In contrast, recruitment of p300 was unaffected by either siA or siD (Fig. 5b). Similar data were observed in the absence of ionizing radiation (data not shown). Neither siA nor siD decreased the level of TLS protein (Supplementary Fig. 15c). These data suggest that ncRNA$_{CCND1}$s combinatorially or cooperatively cause repression of the *CCND1* transcription unit.

Real-time PCR studies revealed the existence of ncRNA$_{CCND1}$s (D and A, but not C) in both high-molecular-mass and low-molecular-mass fractions (Supplementary Fig. 16a; data not shown). RNA oligonucleotides corresponding to the ncRNA$_{CCND1}$-'expressing' regions (for example −454s and −341a) were capable of binding to TLS and inhibiting the HAT function of p300 (Fig. 5c and Supplementary Fig. 16b, c). In contrast, a different series of RNA oligonucleotides evaluated, including oligonucleotides based on the *β-actin* mRNA sequence and the ncRNA$_{CCND1}$-'non-expressing' region C (−764a), were unable to bind to TLS (Supplementary Fig. 16d) or inhibit the HAT functions of p300 (Fig. 5c). Moreover, siA and siD, but not siC or siF, resulted in an increase in histone acetylation (AcH3-K9K14) on the *CCND1* promoter (Fig. 5d).

We suggest a model in which ncRNAs serve as molecular 'ligands' for a specific RNA-binding protein, namely TLS, causing an allosteric effect to release it from an inactive conformation. This in turn permits gene-specific TLS–CBP/p300 interactions resulting in the inhibition of HAT functions of CBP and p300 and the repression of transcription (Fig. 5e). It is tempting to speculate that other RNA-binding co-regulators exert functional roles on gene transcription by being analogously recruited to the transcription units through gene-specific ncRNAs.

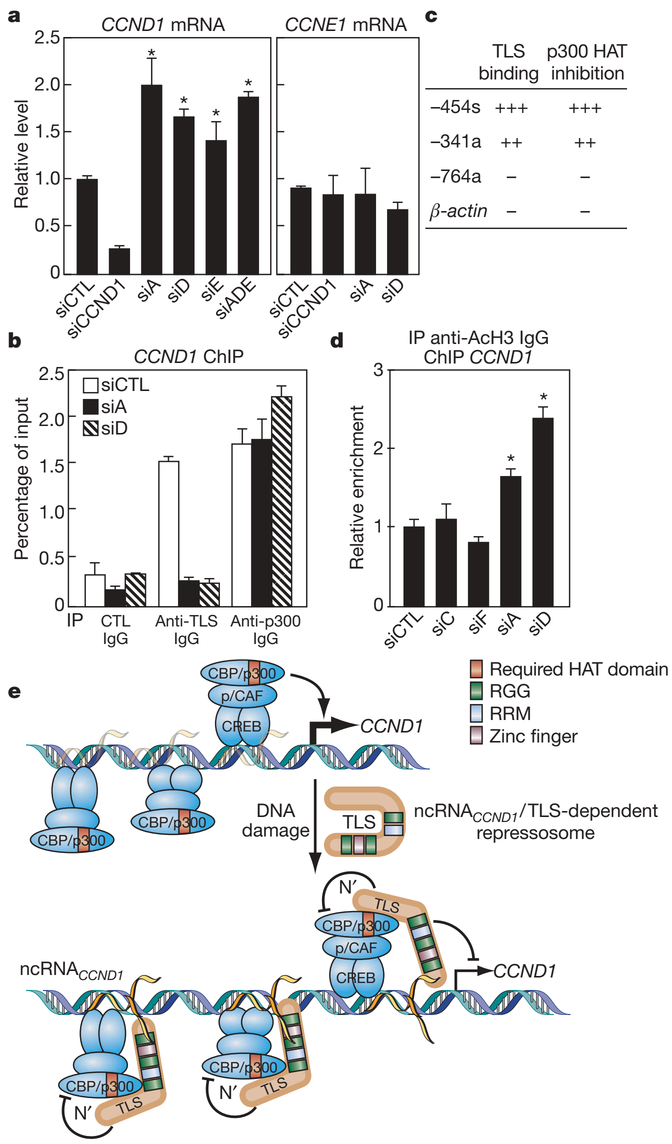


Figure 5 | ncRNA$_{CCND1}$ negatively regulates *CCND1* transcription by recruiting TLS to the *CCND1* promoter. a, Expression levels of *CCND1* and *CCNE1* in the presence of siRNA targeting ncRNA$_{CCND1}$_A (siA), ncRNA$_{CCND1}$_D (siD) or ncRNA$_{CCND1}$_E (siE), or by cotransfection with these siRNAs (siADE), or targeting the *CCND1*-coding region (si*CCND1*). Asterisk, $P < 0.01$ compared with control siRNA (siCTL); $n = 6$. **b**, ChIP of TLS and p300 on the *CCND1* promoter in the presence of siA or siD on treatment with ionizing radiation. **c**, RNA oligonucleotides tested for TLS binding and p300 HAT inhibition. **d**, ChIP of AcH3-K9K14 on the *CCND1* promoter in the presence of indicated siRNAs. Asterisk, $P < 0.05$; $n = 3$. Error bars in **a**, **b** and **d** indicate s.e.m. **e**, Model.

METHODS SUMMARY

RAW264.7 and HeLa cells were maintained in DMEM (Gibco) supplemented with 10% fetal calf serum (FCS; Gemini). Plasmids and siRNAs were transfected with Lipofectamine 2000 (Invitrogen) as directed. Specific antibodies were obtained from BD Biosciences (anti-TLS), Santa Cruz Biotechnology (anti-CBP and anti-p300) and Millipore (anti-acetylated histone H3).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 November 2007; accepted 9 April 2008.
Published online 28 May 2008.

1. Kapranov, P., Willingham, A. T. & Gingeras, T. R. Genome-wide transcription and the implications for genomic organization. *Nature Rev. Genet.* **8**, 413–423 (2007).
2. Bernstein, E. & Allis, C. D. RNA meets chromatin. *Genes Dev.* **19**, 1635–1655 (2005).
3. Carninci, P. *et al.* The transcriptional landscape of the mammalian genome. *Science* **309**, 1559–1563 (2005).
4. Bertone, P. *et al.* Global identification of human transcribed sequences with genome tiling arrays. *Science* **306**, 2242–2246 (2004).

5. Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15** (Spec. Iss. 1) R17–R29 (2006).
6. Martianov, I., Ramadass, A., Serra Barros, A., Chow, N. & Akoulitchev, A. Repression of the human dihydrofolate reductase gene by a non-coding interfering transcript. *Nature* **445,** 666–670 (2007).
7. Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129,** 1311–1323 (2007).
8. Feng, J. *et al.* The Evf-2 noncoding RNA is transcribed from the Dlx-5/6 ultraconserved region and functions as a Dlx-2 transcriptional coactivator. *Genes Dev.* **20,** 1470–1484 (2006).
9. Petruk, S. *et al.* Transcription of *bxd* noncoding RNAs promoted by Trithorax represses *Ubx* in *cis* by transcriptional interference. *Cell* **127,** 1209–1221 (2006).
10. Sanchez-Elsner, T., Gou, D., Kremmer, E. & Sauer, F. Noncoding RNAs of trithorax response elements recruit *Drosophila* Ash1 to Ultrabithorax. *Science* **311,** 1118–1123 (2006).
11. Lanz, R. B. *et al.* A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell* **97,** 17–27 (1999).
12. O'Neill, M. J. The influence of non-coding RNAs on allele-specific gene expression in mammals. *Hum. Mol. Genet.* **14** (Spec. Iss. 1) R113–R120 (2005).
13. Rosenfeld, M. G., Lunyak, V. V. & Glass, C. K. Sensors and signals: a coactivator/corepressor/epigenetic code for integrating signal-dependent programs of transcriptional response. *Genes Dev.* **20,** 1405–1428 (2006).
14. McKenna, N. J. & O'Malley, B. W. Combinatorial control of gene expression by nuclear receptors and coregulators. *Cell* **108,** 465–474 (2002).
15. Seo, S. B. *et al.* Regulation of histone acetylation and transcription by INHAT, a human cellular complex containing the set oncoprotein. *Cell* **104,** 119–130 (2001).
16. Sebastiaan Winkler, G. *et al.* Isolation and mass spectrometry of transcription factor complexes. *Methods* **26,** 260–269 (2002).
17. Uranishi, H. *et al.* Involvement of the pro-oncoprotein TLS (translocated in liposarcoma) in nuclear factor-κB p65-mediated transcription as a coactivator. *J. Biol. Chem.* **276,** 13395–13401 (2001).
18. Yang, L., Embree, L. J., Tsai, S. & Hickstein, D. D. Oncoprotein TLS interacts with serine-arginine proteins involved in RNA splicing. *J. Biol. Chem.* **273,** 27761–27764 (1998).
19. Hicks, G. G. *et al.* Fus deficiency in mice results in defective B-lymphocyte development and activation, high levels of chromosomal instability and perinatal death. *Nature Genet.* **24,** 175–179 (2000).
20. Kuroda, M. *et al.* Male sterility and enhanced radiation sensitivity in TLS$^{-/-}$ mice. *EMBO J.* **19,** 453–462 (2000).
21. Baechtold, H. *et al.* Human 75-kDa DNA-pairing protein is identical to the pro-oncoprotein TLS/FUS and is able to promote D-loop formation. *J. Biol. Chem.* **274,** 34337–34342 (1999).
22. Bertrand, P., Akhmedov, A. T., Delacote, F., Durrbach, A. & Lopez, B. S. Human POMp75 is identified as the pro-oncoprotein TLS/FUS: both POMp75 and POMp100 DNA homologous pairing activities are associated to cell proliferation. *Oncogene* **18,** 4515–4521 (1999).
23. Ron, D. TLS-CHOP and the role of RNA-binding proteins in oncogenic transformation. *Curr. Top. Microbiol. Immunol.* **220,** 131–142 (1997).
24. Kurokawa, R. *et al.* Differential use of CREB binding protein-coactivator complexes. *Science* **279,** 700–703 (1998).
25. Lerga, A. *et al.* Identification of an RNA binding specificity for the potential splicing factor TLS. *J. Biol. Chem.* **276,** 6807–6816 (2001).
26. Miyakawa, Y. & Matsushime, H. Rapid downregulation of cyclin D1 mRNA and protein levels by ultraviolet irradiation in murine macrophage cells. *Biochem. Biophys. Res. Commun.* **284,** 71–76 (2001).
27. Impey, S. *et al.* Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. *Cell* **119,** 1041–1054 (2004).
28. Murata, T. *et al.* Defect of histone acetyltransferase activity of the nuclear transcriptional coactivator CBP in Rubinstein–Taybi syndrome. *Hum. Mol. Genet.* **10,** 1071–1076 (2001).
29. Agami, R. & Bernards, R. Distinct initiation and maintenance mechanisms cooperate to induce G1 cell cycle arrest in response to DNA damage. *Cell* **102,** 55–66 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. Gettings for help with mass spectrometric analysis; M. Hiramatsu, W. Sato and C. Nelson for technical assistance; A. Matsushita, M. Matsubara and T. Oyoshi for discussion; and J. Hightower and M. Fisher for figure and manuscript preparation. This work was supported by the Fujisawa Foundation, the Takeda Science Foundation, the Naito Foundation, Sankyo Foundation Life Science, and grants-in-aid (nos 17054036 and 18055029) from the Ministry of Education, Culture, Sports, Science, and Technology in Japan to R.K., by National Institutes of Health grants CA52599 and HL59694 to C.K.G., by National Cancer Institute Cancer Center Support grant P30 CA08748 to P.T., by NS34934, DK39949 and CA097134 to M.G.R., and by DK074868 to C.K.G. and M.G.R. M.G.R. is a Howard Hughes Medical Institute investigator.

Author Information Reprints and permissions information is available at www.nature.com/reprints. Correspondence and requests for materials should be addressed to R.K. (rkurokaw@saitama-med.ac.jp), C.K.G. (ckg@ucsd.edu) or M.G.R. (mgr@ucsd.edu).

METHODS

Materials and reagents. Antibodies were obtained from Santa Cruz Biotechnology (anti-p/CAF, anti-TAFII68 and anti-EWS), Upstate Biotechnology (anti-TIP60) and Synaptic Systems (anti-cap). siRNAs were obtained from Qiagen: siA, 5′-GGCGCCUCAGGGAUGGCUU-3′; siD, 5′-AAUUCAGUCCCAGGGCAAA-3′; siE, 5′-GACCCGGAAUAUUAGUAAU-3′; siC, 5′-GGCUAGAAGGACAAGAUGA-3′; siF, 5′-GAGUGGGCGAGCCCUCUUUA-3′; si5′UTR, 5′-GGACUUUGCAACUUCAACA-3′; si*CCND1*, SI02654547; siCTL, 5′-AAUUCUCCGAACGUGUCAC-3′; si*TLS*, 5′-CAGAGUUACAGUGGUUAUG-3′ and 5′-UUCUCUGGGAAUCCUAUUA-3′.

HAT assays. HeLa extracts, histones (Sigma) or mononucleosomes (from HeLa cells) and [^{14}C]acetyl-CoA were incubated with baculovirus-expressed CBP in solution HAT assays as described[30]. Pull-down HAT assays were performed by capturing baculovirus-expressed, Flag-tagged CBP on anti-Flag agarose beads (Sigma). Beads were incubated with HeLa extracts for 1 h, washed three times with HAT assay buffer, and then incubated with histones and [^{14}C]acetyl-CoA. CBP and histones were subsequently resolved by SDS–PAGE and acetylation was detected by autoradiography.

Biochemical purification and protein identification. HeLa nuclear extracts were dialysed against 0.1 M NaCl containing dialysis buffer (20 mM HEPES pH 7.9, 0.2 mM EDTA, 0.5 mM dithiothreitol (DTT)), applied to a 500-ml column of Sephacryl S-300, equilibrated, and fractionated into 43 fractions, which were analysed with HAT assay. Fractions with inhibitory activity were further incubated with baculovirus-expressed Flag-tagged CBP bound anti-Flag agarose beads and extracted with 0.3 M NaCl extraction buffer and separated by SDS–PAGE. The protein bands were analysed by matrix-assisted laser desorption/ionization reflectron time-of-flight mass spectrometry (MALDI-re-TOF MS) (UltraFlex TOF/TOF; Bruker) as described[16]. Selected peptide ions (*m*/*z*) were taken to search a 'non-redundant' human protein database (National Center for Biotechnology Information) to identify the proteins.

Gel shift assays. [^{32}P]RNA or DNA oligonucleotides (200,000 c.p.m.) were heated at 95 °C for 2 min and immediately placed on ice. RNA and its cDNA oligonucleotides were heated at 95 °C for 2 min, and annealed down to 25 °C. The probes were then incubated for 15 min at 25 °C in reaction buffer containing baculovirus-expressed TLS, 10 mM Tris-HCl pH 7.5, 5% glycerol, 10 mM EDTA, 1 mM DTT and 5 μg of yeast tRNA. The samples were then analysed on a 6% PAGE gel. The gel was dried and analysed by autoradiography. The RNA and DNA oligonucleotide sequences were as follows: GGUG oligonucleotide, 5′-UUGUAUUUUGAGCUAGUUU<u>GGUG</u>AC-3′; CCUC oligonucleotide, 5′-UUGUAUUUUGAGCUAGUUU<u>CCUC</u>AC-3′; −454s (or RNA in Fig. 4g), 5′-UCUGCCGGCUUGGAUAUGG<u>GGUG</u>UC-3′; −341a, 5′-CCCGGGAUUUAGGGGGUGA<u>GGUG</u>GA-3′; −764a, 5′-UCCAGCAGCAGCCCAAGAU<u>GGUG</u>GC-3′; *β-actin*, 5′-UGGCAUCGUGAUGGACUCC<u>GGUG</u>AC-3′; DNA, 5′-GACACCCCATATCCAAGCCGGCAGA-3′.

RNA extraction and real-time PCR. HeLa cells were lysed in RSB-100 buffer (100 mM Tris-HCl pH 7.4, 100 mM NaCl, 2.5 mM $MgCl_2$, 40 μg ml^{-1} digitonin) followed by centrifugation at 2,000*g* for 8 min. The supernatant fraction was collected as cytosolic fraction. The cell pellet was then resuspended in RSB-100 containing 0.5% Triton X-100 (RSB-100T). After centrifugation at 2,000*g* for 8 min, the supernatant was collected as nuclear fraction. The resulting cell pellet was resuspended in RSB-100T and sonicated (Fisher Sonic Dismembrator, Model 300). The soluble DNA-bound RNA fraction was collected after centrifugation at 4,000*g* for 15 min. RNA was extracted with Trizol (Invitrogen) and treated with RNase-free DNase I (DNA-free; Ambion). Reverse transcription (RT) was performed with a random hexamer or gene-specific primer. Reaction without transcriptase was performed as a no-RT control. Real-time PCR was performed with the Mx3000P (Stratagene).

RNase A, micrococcal nuclease (MNase), DNase I, RNase H and RNase T1 treatment. Whole-cell extracts of GST proteins were treated with RNase A (25 μg per 50 μl; Sigma), and incubated on ice for 20 min. GST–TLS in whole cell extracts was sequentially treated with 10 μg of micrococcal nuclease (Roche) in 100 mM sodium glycine (pH 8.6) and 10 mM $CaCl_2$ at 37 °C for 4 min, 0 °C for 1 min, and 25 °C for 20 min, and terminated by the addition of 10 mM EGTA, followed with or without incubation with RNA oligonucleotides at 100 pmol per 20 μl. GST–TLS was treated for 30 min at 37 °C with DNase I (1 μg per 50 μl) in 50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 50 μg ml^{-1} BSA. For co-immunoprecipitation and RT-real time PCR, cell fractionation extracts containing the DNA-bound RNA were obtained as described before and for 30 min treated with 50 ng μl^{-1} RNase A (Sigma), RNase H (1 U per 10 μl; Invitrogen), or RNase T1 (1 U per 10 μl; Ambion) at 25 °C.

Chromatin immunoprecipitation (ChIP). Cells were crosslinked with 1% formaldehyde and stopped with glycine solution (125 mM). The cells were then sequentially washed in ice-cold buffer I (0.25% Triton X-100, 10 mM EDTA, 0.5 mM EGTA, 10 mM HEPES pH 6.5) and buffer II (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM HEPES pH 6.5). Cell pellets were resuspended in lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl pH 8.1, 1 × protease inhibitor cocktail) and sonicated. The soluble chromatin was then diluted in dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl pH 8.1, 1 × protease inhibitor cocktail). Protein A/G–sepharose beads were added and incubated for 1 h at 4 °C for preclearing. Specific antibody was added to the supernatant and incubated at 4 °C. The next day, Protein A/G–sepharose beads were added and incubated for 2 h at 4 °C. Beads were harvested by centrifugation and washed sequentially in TSE I buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 150 mM NaCl), TSE II buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.1, 500 mM NaCl), buffer III (0.25 M LiCl, 1% Nonidet P40, 1 mM EDTA, 10 mM Tris-HCl pH 8.1) and TE buffer. DNA fragments were eluted overnight in 1% SDS, 0.1 M $NaHCO_3$ at 65 °C and purified with a QIAquick Spin Kit (Qiagen).

RNA immunoprecipitation assay. Whole-cell extracts were obtained in NETN buffer (125 mM NaCl, 1 mM EDTA, 20 mM Tris-HCl pH 8.1, 0.5% Nonidet P40, 10% glycerol, 1 × protease inhibitor cocktail) without crosslinking, followed by sonication and preclearing as described for ChIP assay. Conjugated antibody/protein A/G–sepharose beads were pretreated with RNase inhibitor and then added for a further incubation at 4 °C overnight. Beads were then washed at least six times for 10 min each at 4 °C in NETN buffer. Bound RNA was then eluted from the beads by directly adding Trizol (Invitrogen) to the beads, followed by RNA extraction and RT-real time PCR as described previously.

30. Korzus, E. *et al.* Transcription factor-specific requirements for coactivators and their acetyltransferase functions. *Science* **279**, 703–707 (1998).

naturejobs

THE CAREERS MAGAZINE FOR SCIENTISTS

A survey of 109 UK universities says that academic life is getting sweeter. The study, conducted by Hampshire-based Incomes Data Services for the Universities and Colleges Employers Association, says that academics are contracted to work fewer hours, have more holiday time, and better child-care policies, pension plans and sick pay than elsewhere in the public and private sectors. A related report also says that academic pay has improved in recent years, with a 30% increase between 2002 and 2007 putting the average salary at £42,588 (US$85,000) in 2007.

This sweet spin on academic employment has left a sour taste in some mouths, however. In the *Times Higher Education Supplement* (*THES*) in London, Sally Hunt, general secretary of the University and College Union, said that the report used data on contracted hours rather than actual time worked, thus painting a "misleading picture". Other surveys show that academics work some 50+ hours a week, rather than the 35 hours a week the survey provided as the contracted average. In a related blog, several *THES* readers agreed, saying that they put in far more than the contracted weekly hours and often forgo holiday time in order to catch up. And when they do take leave, they often take work with them. As far as salaries go, Hunt agrees that academics' pay has increased — but she says that it still lags behind equivalent positions in the private sector.

Is this just envy speaking? Perhaps, given that academic positions are highly sought after. Some might point out that academics don't have to put in the extra hours — in return, researchers would argue that unpaid overtime is mandatory if you are going to get published and win grants.

The report did find some negative spots. It warned of annual budget uncertainties that affect research staff and the proliferation of short-term contracts. These are the real factors that need addressing. Otherwise, there will be fewer academics in permanent positions left to quibble about their hours and more researchers working longer and harder to fight for less funding. ■

Paul Smaglik moderates the Naturejobs Nature Network career-advice forum.

CONTACTS
Editor: Gene Russo

European Head Office, London
The Macmillan Building,
4 Crinan Street, London N1 9XW, UK
Tel: +44 (0) 20 7843 4961
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

European Sales Manager:
Andy Douglas (4975)
e-mail: a.douglas@nature.com
Business Development Manager:
Amelie Pequignot (4974)
e-mail: a.pequignot@nature.com
***Nature*events:**
Claudia Paulsen Young (+44 (0) 20 7014 4015)
e-mail: c.paulsenyoung@nature.com
France/Switzerland/Belgium:
Muriel Lestringuez (4994)
Southwest UK/RoW: Nils Moeller (4953)
Scandinavia/Spain/Portugal/Italy:
Evelina Rubio-Hakansson (4973)
Northeast UK/Ireland:
Matthew Ward (+44 (0) 20 7014 4059)
North Germany/The Netherlands:
Reya Silao (4970)
South Germany/Austria:
Hildi Rowland (+44 (0) 20 7014 4084)

Advertising Production Manager:
Stephen Russell
To send materials use London address above.
Tel: +44 (0) 20 7843 4816
Fax: +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com
***Naturejobs* web development:** Tom Hancock
***Naturejobs* online production:** Dennis Chu

US Head Office, New York
75 Varick Street, 9th Floor,
New York, NY 10013-1917
Tel: +1 800 989 7718
Fax: +1 800 989 7103
e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

India
Vikas Chawla (+91 1242881057)
e-mail: v.chawla@nature.com

Japan Head Office, Tokyo
Chiyoda Building, 2-37 Ichigayatamachi,
Shinjuku-ku, Tokyo 162-0843
Tel: +81 3 3267 8751
Fax: +81 3 3267 8746

Asia-Pacific Sales Manager:
Ayako Watanabe (+81 3 3267 8765)
e-mail: a.watanabe@natureasia.com
Business Development Manager, Greater China/Singapore:
Gloria To (+852 2811 7191)
e-mail: g.to@natureasia.com

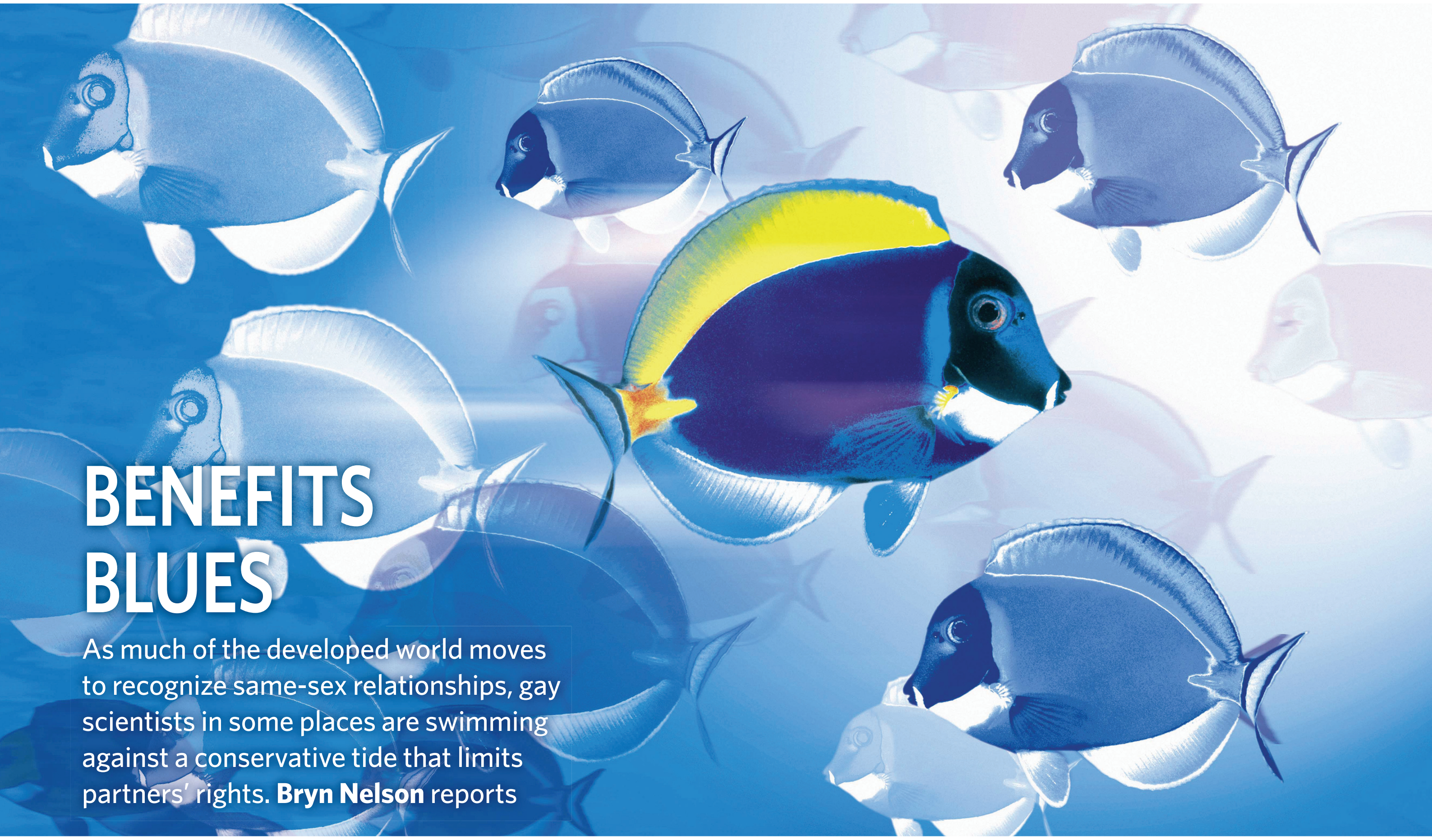


IMAGE SOURCE/CORBIS

BENEFITS BLUES

As much of the developed world moves to recognize same-sex relationships, gay scientists in some places are swimming against a conservative tide that limits partners' rights. **Bryn Nelson** reports

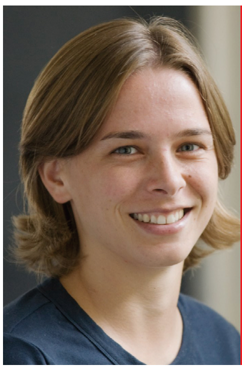
February was a bittersweet month for Michael Falk. The associate professor of materials science and engineering at the University of Michigan in Ann Arbor was honoured as educator of the year by the National Organization of Gay and Lesbian Scientists and Technical Professionals (NOGLSTP) only days before beginning a move to Johns Hopkins University in Baltimore.

The main motive for Falk's move was neither money nor prestige, but a Michigan constitutional amendment forbidding public universities and agencies to grant benefits relating to an employee's unmarried 'domestic partner'.

"It wasn't like I was leaving because I didn't have enough lab space or was seeking more institutional support," Falk says. "It was like, 'Can you change the state constitution?'"

For many lesbian, gay, bisexual and transgender academics around the world, issues such as job security and peer support are top concerns. But a growing number are moving long distances, or out of academia altogether, on the basis of benefits often beyond an employer's power to grant.

The shift has created new tensions between public universities and their funding sources and has prompted fears, especially in the United States, of a brain drain to more accommodating places. Even as New York and California are dramatically widening their recognition of gay marriage, liberal university towns in the Midwest are bidding farewell to top talent after recent state amendments blocked even straight domestic partnerships. Job-seekers have weeded out prospective employers for the same reason. Scientists from European and other countries are incorporating gay-friendly immigration laws into their decisions about where to work abroad.



"The University sponsorship scheme ensures that de-facto partners of either gender can receive visas."— Lisa Harvey-Smith

SCHOOL OF PHYSICS, UNIVERSITY OF SYDNEY

Hodgepodge of US laws

The legal picture is particularly muddled in the United States, where job-seekers have to navigate a hodgepodge of local, state and federal laws covering partnership recognition, parenting rights and job discrimination. Last year, Michigan's court of appeals ruled unanimously that the wording of a constitutional amendment, passed in 2004, barred state-funded institutions from extending insurance coverage to domestic partners of employees (a decision upheld by the state supreme court in May). The appeals court ruling convinced Falk and his partner to consider moving away. The University of Michigan and several other state-funded campuses have since set up a category for "other eligible individuals" living with employees. The revision has allayed some concerns about insurance policies and let qualified dual-sex couples claim the same benefits. But opponents are considering a court challenge on the grounds that the new category serves the same purpose as the old one.

Other states have seen similar cases. Robert Carpick, an associate professor of mechanical engineering and applied mechanics at the University of Pennsylvania in Philadelphia, says that a 2006 ban on same-sex marriage in Wisconsin, also interpreted as prohibiting

domestic-partnership rights, influenced his decision to leave the University of Wisconsin at Madison. "It was unacceptable for me to be treated differently from my colleagues," says Carpick, a Canadian citizen who legally married his partner in 2003 in Ontario. Now co-director of the Nanotechnology Institute at the University of Pennsylvania, he says he took this position for many reasons. Nevertheless, he would not have considered an offer that lacked domestic-partner benefits. "That was a necessary condition," he says.

In 2006, Arizona voters narrowly defeated a constitutional amendment that many believed would have barred domestic-partner benefits for both gay and straight couples. Earlier this year, the state adopted a policy that grants such benefits to public employees. "In our view, it's the right thing to do and it gives us a statewide policy," says Arizona State University spokesman Virgil Renzulli. "We think, obviously, we're in a national and international market for faculty, and that it's important that we offer competitive packages." The universities of Louisville and Kentucky cite a similar rationale for offering domestic-partner benefits in 2006. Earlier this year, state lawmakers tried unsuccessfully to scuttle them with a new law.

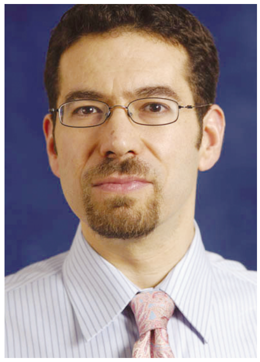
Trying to stay competitive

In Florida, public institutions cannot use state money for extending health benefits to employees' domestic partners. The University of Florida-Gainesville, which has provided such coverage since 2006, pays through contracts, federal grants and auxiliary funds. Kyle Cavanaugh, vice-president of human resources, says that most of the university's US peer institutions offered similar benefits and that, to remain competitive, it couldn't afford not to.

Yet the university may lose its say in November, when a ballot could replace Florida's prohibition on same-sex marriage with a constitutional ban, stating "no other legal union that is treated as marriage or the substantial equivalent thereof shall be valid or recognized". Opponents say the wording could be interpreted to bar any domestic-partnership recognition. "We would hope that we would not see a situation similar to what happened in Michigan," says Cavanaugh.

Sophie Parker, who was until recently a postdoc in ecosystems and ecology at the University of California, Santa Barbara, left academia rather than leave California's progressive laws. Parker abandoned a promising job-interview process in Chicago, largely over concerns that an Illinois law banning same-sex marriage would be strengthened to prohibit any partnership recognition, which could interfere with her plans to have children. Instead, she stayed in California and took a job at the Los Angeles office of the Nature Conservancy. A high-profile state supreme court ruling has since opened the doors to same-sex marriage in California, although voters could reverse this with a constitutional amendment later this year.

No estimates exist for the number of gay, lesbian, bisexual and transgender researchers in academia, much less for those facing similar dilemmas. Cases of career change as a result of constitutional barriers are probably under-reported because most affected



Moving out: Sophie Parker (top) gave up academia; Michael Falk left Michigan.

A. MARTIN

"It was unacceptable for me to be treated differently from my colleagues." — Robert Carpick

JEROME JIM

researchers are too closeted to raise a fuss, says Rochelle Diamond, chairwoman of the NOGLSTP. "If they're research scientists, they have grants, and they fear what their peer-review committees would think of them," she says. "And there's a lot of prejudice in the ranks for tenure review."

Easier in England and Australia

The picture is far different at some other institutions around the world. Lisa Harvey-Smith, a postdoctoral fellow in radio polarimetry at the University of Sydney in Australia, says she has had no problems being 'out', either at the university or back in her native England, where she studied at the University of Manchester's Jodrell Bank Observatory. She praises the "fantastic" stance taken by Australia and the University of Sydney in recognizing her relationship.

"It is generally quite difficult to get a working visa for Australia but the University [of Sydney] sponsorship scheme ensures that de-facto [unmarried] partners of either gender can receive visas with full work and study rights for the duration of your employment contract," Harvey-Smith says. Not so in the United States, where same-sex partners can tag on to a foreign researcher's visa but cannot work. Unless a change is in the offing, Harvey-Smith says, "archaic immigration laws" will prevent her from working there.

Despite South Africa's racist history and a recent spate of violence against poor immigrants, the country has a rather gay-friendly constitution. Astronomer Patrick Woudt says that South Africa's tolerant work environment eased his career move to the University of Cape Town. A Dutch citizen who earned his PhD in Cape Town, Woudt says he has been completely open about his sexual orientation since returning in 2000 as a postdoc. "I haven't experienced anything negative in all the time I have been at the University of Cape Town," says Woudt, now a senior lecturer in astronomy. After he met his South African husband seven years ago, they were married in the Netherlands in 2004. South Africa now also recognizes same-sex marriages.

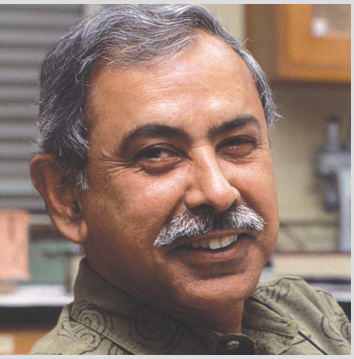
Places such as India are more challenging. One scientist who trained in New York says she and most other gay Indian scientists remain closeted. The researcher, who asked not to be identified for fear of reprisal, says the obstacles she faces as a female scientist would only be compounded were she to be openly gay. She has tried to combat her isolation by discreetly approaching other gay researchers about forming a support network. "It's important for younger people to have role models," she says. So far, however, her efforts have been in vain.

Diamond says that NOGLSTP has tried to fill the mentorship gap by establishing an Internet-based programme affiliated with MentorNet. Scientists and engineers can mentor students, postdocs and young professionals one-to-one, offering guidance on being 'out' at work as well as on other topics that could prove awkward for advisers. "Students have questions that most career counsellors can't answer, like 'Do I put the picture of my honey on my desk?'" Diamond says. Increasingly, the replies could help determine where they work next. ■

Bryn Nelson is a freelance science writer based in Seattle, Washington.

MOVERS

Sudhansu K. Dey, director of Division of Reproductive Sciences, Cincinnati Children's Research Foundation



2002–08 Director, Division of Reproductive and Developmental Biology, Vanderbilt University Medical Center, Nashville, Tennessee
1973–2002 Postdoctoral fellow to professor, University of Kansas Medical Center, Kansas City, Kansas

After earning a PhD in India, Dey landed a postdoc in the United States, But when he arrived in Kansas City, Kansas, in 1972 with eight dollars in his pocket, he had nowhere to stay and no one greeted him. His mentor's number was unlisted. Eventually, with the help of the university's vice chancellor, he found some lodging.

"That was a pretty scary 24 hours," Dey says. It was an inauspicious beginning to a productive research career; Dey ended up staying at the university nearly 30 years and publishing hundreds of papers on reproductive biology, an ambition inspired by watching his parents struggle to provide for their 12 children on the outskirts of Calcutta.

But things got worse in Kansas City before they got better. On his own initiative, Dey studied whether preimplantation embryos produce steroid hormones, which are critical for embryo implantation. But it was his mentor, not Dey, who received first authorship — the first of a few such incidents. With thoughts of abandoning his postdoc and the United States, Dey visited the department chair, who offered Dey a job independent of his former mentor, with one year to get his own funding.

"That was the best mentoring I ever had to steer through my career path," says Dey. He literally knocked on doors and formed collaborations at Kansas City "out of necessity". The strategy worked. Dey eventually earned his own first-author paper and landed National Institutes of Health funding.

His basic reproductive biology research encompasses embryo implantation and several forms of cancer, and some stem-cell work. "I am quite aggressive in terms of collaboration," Dey says. "I just pick up the phone and ask." That approach helped get him recruited at Vanderbilt.

Dey's Vanderbilt enthusiasm was short-lived, owing to limited institutional support for reproductive sciences. Former Vanderbilt colleague Arnie Strauss, now head of Cincinnati Children's Research Foundation, asked Dey to come to the foundation's Cincinnati Children's Hospital Medical Center. Strauss clinched the deal by asking Dey to start a new division of reproductive sciences.

"He's a tremendous collaborator," says Strauss. He calls Dey a "tenacious" investigator, who tackles difficult goals by bringing people together and sharing credit. Dey says he learned his biggest lesson — one he passes on to students, postdocs and junior faculty — on that first day in Kansas City. "You have to be stubborn but generous," he says. ■

Paul Smaglik

BRICKS & MORTAR

Shock physics

The growing field of shock physics has received crucial support from the UK government. Imperial College will get £10 million (US$20 million) over five years to establish a research institute for shock physics: the study of matter at extreme conditions and how materials respond to shock waves, high pressures and temperatures. The funding comes from the UK's Atomic Weapons Establishment (AWE), the private contractor charged by the UK Ministry of Defence to develop and maintain nuclear weapons.

Imperial's new Institute for Shock Physics will consolidate theoretical and computational research by Earth scientists, engineers and energy researchers. They will investigate the physical properties of materials exposed to extreme pressures or temperatures such as those associated with asteroid impact on Earth.

Although no defence research will be carried out at the institute, AWE hydrodynamics scientist David Holder says it will be an important training ground for future employees. The directorship, three faculty positions and 20 PhD fellowships are currently available. As of 2009, a one-year MSc course will be offered.

Interim director Steven Rose says interest in new energy sources such as fusion and in understanding phenomena such as tsunamis are among the factors driving the field. "In the UK, there is yet no single point of focus for these disparate bits of related research," he says. New equipment, such as a gas-gun able to fire a projectile at several kilometres per second to create high-velocity impacts, will help researchers investigate, for example, how aircraft would respond to a meteorite strike.

The additional infrastructure could be key for the field as a whole. "With the Imperial institute, the field of shock physics gets a boost of experimental capabilities," says Yogendra Gupta, director of Washington State University's Institute for Shock Physics and a leader in the field.

Gupta says the recent ability to achieve pico- and nanosecond measures of matter being destroyed at extreme conditions sheds light not only on the fundamental physics necessary to develop the next generation of materials, but on new ways to generate energy. New opportunities should come next year with the completion of the US Department of Energy's National Ignition Facility. It will enable researchers to conduct fusion experiments by squeezing matter with shock waves. ■

Virginia Gewin

POSTDOC JOURNAL

Getting people to talk

Conferences charge me with motivation: they provide opportunities to share data, ideas and enthusiasm. But it is easy to make the excuse that one is "too busy" to attend. It is thus with anticipation and trepidation that I am helping to organize the first annual symposium for students and postdocs at my institute.

A previous postdoc seminar series was run single-handedly by fellow postdoc Claire Canning. To get speakers, she had to send e-mails that were alternately wheedling and forceful. Even within the institute, people seemed wary of sharing unpublished data. One even demanded that we somehow guarantee presented data would not be misappropriated. Still more discouraging was the turnout: fewer than 20 people from an institute of more than 400. The poor response may have been partly due to the lack of involvement and support by lab leaders. A few students told me that their supervisors specifically instructed them to focus on their benchwork rather than participate. Claire's tenacity and valiant effort truly deserves credit, as she kept the seminars going for almost a year.

Despite her experience, I hope the forthcoming symposium will see more enthusiastic participation. This time we have the active support of our institute director. There will be food, drink, and even prizes! Hopefully we will all come away with a stronger sense of community, well-fed in both mind and body. ■

Amanda Goh is a postdoctoral fellow in cell biology under the Agency of Science, Technology and Research in Singapore.

NIH International Research Career Transition Programs

This is an exciting time for biomedical research. New technologies are enhancing our ability to understand the complex interplay between genes and the environment, and new findings are providing novel insights into human disease. Research from the bench to the bedside – and back again – is becoming a reality. The NIH is committed to playing a leading role in this enterprise, harnessing the power of all the sciences - biology, chemistry, physics, engineering, computer science, bioinformatics, and the behavioural/social sciences - *to improve human health globally*.

The NIH Intramural Research Program (IRP) conducts basic, translational, epidemiological, and clinical biomedical research in NIH facilities on the main campus in Bethesda, MD, and other campuses throughout the U.S. while providing training opportunities for more than 6,000 young scientists, including more than 3700 postdoctoral fellows. Trainees from around the world make key contributions to the success of the NIH IRP; almost two-thirds of our postdocs are foreign nationals. In turn, the NIH considers itself responsible for ensuring the career success of its international postdocs and working with them to enhance global collaboration.

Over the past year, working together with the NIH Institutes and Centers, the NIH Office of Intramural Training & Education (OITE) has dramatically enhanced the services available to all trainees in the IRP. We have established and staffed a Career Services Center and an Office of Postdoctoral Services. Series of workshops and activities targeting the academic and industry job search processes, written and oral communication skills, teaching, and leadership are either under development or being offered. We have placed particular emphasis on opportunities for fellows from other countries to improve their spoken English.

An especially difficult period in the scientific career trajectory, in the U.S. and abroad, is the transition from a mentored position to research independence. To facilitate this transition, the NIH recently implemented its Pathway to Independence (K99/R00) award, which assists individuals of all nationalities, who are conducting postdoctoral research in the U.S., with the move to a faculty position *at a U.S. institution*. Since 2005, the NIH, through its Fogarty International Center, has also worked with countries, regions, and government science agencies around the world to establish a parallel group of programs, the **"NIH International Research Career Transition Programs"**, that enable recent doctoral degree recipients from the participating countries to conduct postdoctoral research at the NIH and then return to positions as independent researchers *in their home countries*. Although the programs vary from country to country, all have the following elements in common.

- The program consists of two phases: Phase I is a postdoctoral research training experience completed at the NIH; in Phase II the fellow returns to a funded research position in his/her home country.
- Program applicants are selected through a competitive application process, with the review managed by the home country.
- Individuals selected for the program identify a researcher at the NIH who is willing to provide funding for Phase I.
- During the fellow's stay at the NIH, the home country funds regular trips home to enable the fellow to network and maintain contact with the home scientific community.
- Upon completion of Phase I, the fellow returns to a funded, independent research position of several years' duration (Phase II).

Other details of the programs, such as eligibility criteria, number of application cycles per year, and application deadlines, depend on the home country/region.

At present, the NIH maintains active partnership agreements with

- Germany [Deutsche Forschungsgemeinschaft (DFG)]
- France [Institut National de la Santé et de la Recherche Médicale (INSERM) and Centre National de la Recherche Scientifique (CNRS)]
- Flanders [Research Foundation - Flanders (FWO)]
- Chile [Comisión Nacional de Investigación Cientifica Y Tecnológica (CONICYT)]
- Institutions in the Lombardy region of Italy (Regione Lombardia)
- The Andalusian Regional Ministry of Health in Spain [Consejeria de Salud de la Junta de Andalucia (CSJA)]

Information on all of these programs can be found at http://www.training.nih.gov/postdoctoral/international.asp.

NIH International Research Career Transition Programs, partnerships between the NIH and the international research community, represent a unique opportunity to build relationships between NIH investigators and the best young scientists world-wide. It is our hope that program participants will return to satisfying careers in their home countries where they will enhance the research environment while maintaining collaborative ties with their NIH hosts.

Patricia M. Sokolove, PhD
Deputy Director, Office of Intramural Training & Education
sokolovp@mail.nih.gov

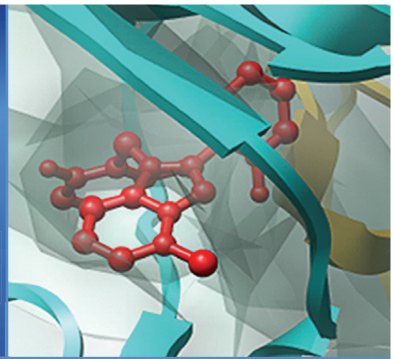
National Cancer Institute

The Center for Cancer Research presents:

Chemical Insights into Biological Processes

Organized by the Chemistry and Structural Biology Faculty

August 15–16, 2008
Hood College, Frederick, MD



Invited Speakers Include:

Gregory L. Verdine, Ph.D.
Harvard University

Carolyn Bertozzi, Ph.D.
University of California, Berkeley

Benjamin F. Cravatt, III, Ph.D.
Skaggs Institute for Chemical Biology, Scripps Research Institute

Jon Clardy, Ph.D.
Harvard Medical School

Peter G. Schultz, Ph.D.
Scripps Research Institute

Angela Belcher, Ph.D.
Massachusetts Institute of Technology

James R. Heath, Ph.D.
California Institute of Technology

Laura L. Kiessling, Ph.D.
University of Wisconsin-Madison

Lawrence J. Marnett, Ph.D.
Vanderbilt Institute of Chemical Biology

Milan Mrksich, Ph.D.
University of Chicago

Alanna Schepartz, Ph.D.
Yale University

Mike E. Jung, Ph.D.
University of California, Los Angeles

Cynthia J. Burrows, Ph.D.
University of Utah

Moungi G. Bawendi, Ph.D.
Massachusetts Institute of Technology

Space is limited, so advance registration is required by August 1, 2008, at:

www.ncifcrf.gov/events/ccrcbs/2008

The deadline for poster abstract submissions is July 25, 2008.

NATIONAL CANCER INSTITUTE
CENTER FOR CANCER RESEARCH

NIEHS
National Institute of Environmental Health Sciences

Department of Health and Human Services
National Institutes of Health (NIH)
Director, National Institute of Environmental Health Sciences (NIEHS)

THE POSITION: The NIH is seeking exceptional candidates for the position of Director, NIEHS, to provide leadership to one of the preeminent centers for environmentally-related research in the world. The Director, NIEHS, also serves in a dual role as the Director, National Toxicology Program, and in this role reports to the Secretary, Department of Health and Human Services (DHHS). This position offers a unique opportunity for the right individual to provide strong and visionary leadership to an organization dedicated to reducing the burden of human illness and dysfunction from environmental causes by understanding each of these elements and how they interrelate. The Director will manage a high-level complex organization and must demonstrate integrity and fairness, adhering in work and behavior to the highest ethical standards, and upholding the highest standards of scientific research and/or business practices. Applicants must possess an M.D. and/or Ph.D. and have senior-level research experience and knowledge of research programs in one or more scientific areas related to environmental effects on human health and/or toxicology research. They should be known and respected within their profession, both nationally and internationally, as individuals of outstanding scientific competence. Salary is commensurate with experience, and full Federal benefits, including leave, health and life insurance, retirement and savings plan (401K equivalent) will be provided. A detailed vacancy announcement that includes application procedures is available at: **http://www.jobs.nih.gov (under Executive Jobs)**. NIEHS is located in **Research Triangle Park (RTP), North Carolina**. Questions may be addressed to **Ms. Lynnita Jacobs at: SeniorRe@od.nih.gov**. CV and bibliography must be received by **11:59 p.m., Monday, August 4, 2008**.

NW163530R

THE NIH IS DEDICATED TO BUILDING A DIVERSE COMMUNITY IN ITS TRAINING AND EMPLOYMENT PROGRAMS

NATURE|Vol 454|3 July 2008

When Johnny comes marching home

Casualties of war.

Chaz Brenchley

The war was a terror, but it was the peace that terrified. Like the vast bleak mysteries of adulthood — so much space and empty yet, quiet yet, unmarked by anything of ours — after the battlefield that was school. Like school, the war was none of our doing, it had simply happened to us and was always obviously survivable. Now we had somehow to make things happen on our own account, with no idea what would happen if we failed.

The aliens didn't help — but that was the point, rather. They weren't here to help. None of us understood them, quite. Not an invasion or an occupying force — because of course we hadn't lost the war, exactly. They might have been ambassadors or tourists, traders or artisans or thieves. Or weapons inspectors, because if we hadn't quite managed to lose the war we most certainly had not won it.

They might have been weapons. It was very hard to tell, and the government wasn't telling. The government gave a fine impression of knowing very little more than we did.

The aliens didn't really matter, though, day to day. They were a mystery and could stay that way, an expression of faith; we saw very little of them, day to day.

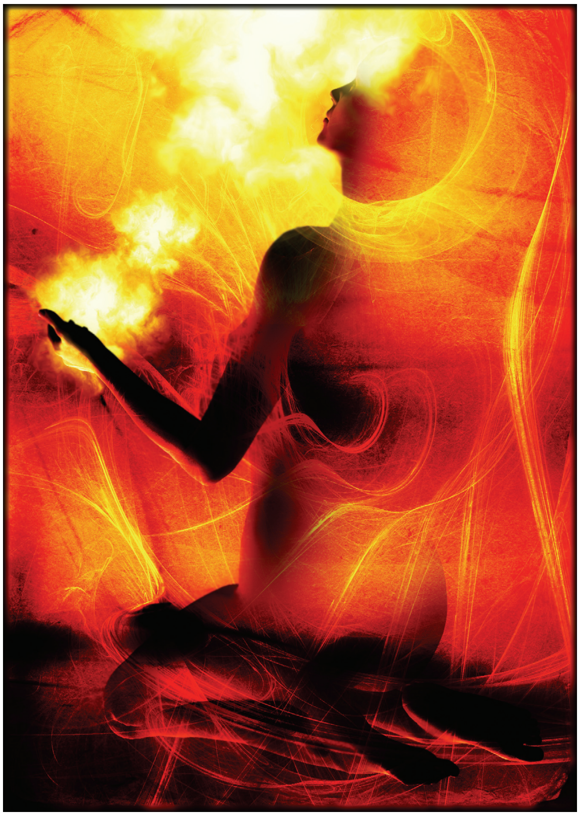
Our own soldiers were another matter. The survivors, that is. They were everywhere, inescapable, returned.

The wounded we could deal with. Some we could even treat. The psychologically scarred, those too; that was old science, albeit faced with new hurts.

But then there were the others, those who came back altered in ways we didn't understand. Enhanced, some people said. Or cursed, or corrupted. Betrayed, or betraying — some called them traitors, spies, dishonoured.

Some called them gods.

I had one who sat below my window, day and night. She didn't beg or steal. She took nothing from the government or anyone else. So far as I could tell, she neither ate nor excreted. Nor slept. Whenever I looked — and sometimes I would spend days, nights, just watching her — she was exactly there, exactly where she had been, sitting cross-legged and playing with light. Running little balls of it up and down her arms, drawing lines of it in the dust, weaving patterns between her hands in a vivid and complex cat's cradle. Perhaps she lived on light now; perhaps she drew nourishment from the glowing air that engulfed her. Perhaps she was beyond feeding, as she was beyond sleep.



JACEY

Some returnees were studied, of course, but there were too many and nowhere to house them; in honesty, nowhere that seemed safe, so many people so strangely changed and all together. Perhaps they were a danger after all?

So they leaked back into the world, and talked or did not talk, and were not understood either way. Not all could play with light. Some could shape sound, make an obscure sculpture of it, a new art that seemed to baffle them as much as us. Some were physically altered; I met a slipskin once, who could never quite be touched, as though he was magnetized to the same polarity as the world around. He was naked, necessarily, and there was a measurable distance between his feet and the floor. It seemed not to inhibit him, and not to be any kind of use.

Generally they seemed to have small use for us, or for any life that we could offer. Their eyes were turned to another horizon. Mostly they were patient, but not all, or not without limits. My own lightweaver: one night she stood and drew herself a framework as high and wide as she could reach, lines of light that burned in the air; then she stepped through it and was gone before those lines had faded.

Some slipskins went too, as though neither gravity nor love could hold them any longer, as though there was nothing they could cling to. They needed no bright gateway; their own condition took them away, at an angle that could be neither described nor recorded.

We looked more closely then at those who remained, those who would let us look. There was no breakthough, more the slow dawn of a consensus. These weren't victims, any more than they were accidents. They were templates, rather. Armatures. Design specs.

There was a phase-shift out there, waiting to be found: an alternate way to be, or to go. No wonder the aliens had proved unreachable; no wonder our war had fizzled out in mutual helplessness, where neither side could entirely reach the other. But these people, salvaged prisoners of that war, they were our guides into crossing the border. Or they could be, as soon as we had solved the equations of light and skin and absence. Monkey see, monkey do: knowing that it was possible, we could do this.

We can do this. Not yet, but we'll get there. And then we'll get — somewhere else. Which is terrifying, true, but probably essential. It is what's waiting for us now. What these people have been altered for: skeleton keys to an uncertain lock. A gesture, a gift, an introduction. A helping hand.

There are still those who call them traitors, fifth-columnists, Trojans within the walls.

I call them Marshalls. With all that that implies. ■

Chaz Brenchley has been a professional writer since he was 18, working mostly in crime, horror, fantasy and science fiction. He claims to live down the dirty end of genre fiction. In fact, he lives in Newcastle upon Tyne.